Evaluating Interfaces for Intelligent Mobile Search

Karen Church, Barry Smyth and Mark. T. Keane
Adaptive Information Cluster,
Smart Media Institute, Department of Computer Science,
University College Dublin, Belfield, Dublin 4, Ireland
{Karen.Church, Barry.Smyth, Mark.Keane}@ucd.ie

ABSTRACT

Recent developments in the mobile phone market have led to a significant increase in the number of users accessing the Mobile Internet. Handsets have been improved to support a diverse range of content types (text, graphics, audio, video etc.), infrastructure investments have delivered improved bandwidth, and changes to billing models offer users much greater value for content. Today large numbers of users are moving away from browsing operator portals and towards off-portal search, leading to a growing need for mobile specific search engine technologies. In this paper we argue that existing mobile search engines are unlikely to offer an adequate service for mobile searchers. Most borrow traditional query-based search and list-based result presentation formats from Web search and as such are not well optimised for the input and display features of mobile devices. For example, many simply attempt to *translate* Web content for the mobile space which is not appropriate. In this paper we evaluate an alternative strategy which replaces the usual result snippet with a more economic alternative that is composed of the keywords used in related queries. We argue that this alternative is better suited to the display characteristics of mobile devices, without compromising the informativeness of result snippets.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.2 [**Information Interfaces and Presentation (e.g.,HCI)**]: User Interfaces

General Terms

Design, Experimentation, Human Factors

Keywords

Mobile Search, Mobile Web, Mobile Internet, Mobile Interfaces, Search Interfaces, User Evaluation

1. INTRODUCTION

New life has been breathed into the Mobile Internet as a result of a combination of significant device, content, infrastructure and billing improvements. The result is that Mobile Internet usage is growing at a significant rate. According to a recent report published by Strategic Analytics, the total number of mobile phone subscribers worldwide approached 2.2 billion at the end of 2005 and looks set to reach 2.5 billion by the end of 2006 [11]. Over 800 million mobile phones were sold in 2005 and as the prices of cellular handsets continue to drop, this figure looks set to rise to 930 million [7] in 2006. The latest statistics also indicate that there has been a significant increase in the number of users accessing the Mobile Web. Ipsos Insight, a market analysis company, have recently published a web study showing that 28% of mobile phone subscribers worldwide have used their phones to browse the Internet, an increase of approx. 3% on the figures released in 2004. Interestingly, this pattern of growth was driven primarily by more mature users (age 35+) indicating that the traditional early adopter group, i.e. young males, no longer dominate Wireless Internet access [17].

Recent trends suggest that users are beginning to move away from the traditional walled garden of the operator portal as they begin to explore the burgeoning off-portal content. A similar effect can be traced back to the early growth of the World-Wide Web as users who had previously been content to browse early portals such as Yahoo, quickly began to explore the greater Web with the help of the latest search engines. And so we might expect mobile search to quickly come to dominate as the primary mode of information access for users, as it has in the World-Wide Web. Certainly there has been significant industry activity in the mobile search space with major players within the search engine industry venturing into the mobile sector. Google and Yahoo have released a number of mobile search solutions including a local search service and an SMS-based search service. Ask Jeeves is currently developing a new wireless search application and America Online (AOL) have recently added enhancements to their mobile search solution. A number of new mobile-specific search services have also come onto the market including Mooobl [1], 4info[2], UpSnap[3] and Technorati Mobile[4].

However, despite this flurry of activity it is our contention that the state of mobile search is not a healthy one. In particular, we believe that the current strategy of simply retrofitting traditional search engine technologies borrowed from the Web for the mobile space is ill conceived. In this paper we argue that such an approach is unlikely to succeed because of the significant and unique challenges presented by the Mobile Internet. We begin with a summary review of the current state of mobile search focusing on an evaluation of 7 existing mobile search engines. The results point to

W4A at WWW2006 23-26 May 2006, Edinburgh, UK

[1]http://www.mooobl.com/
[2]http://www.4info.net/
[3]http://www.upsnap.com/
[4]http://m.technorati.com/

some significant limitations of current approaches and in particular highlight some serious problems when it comes to presenting search results on small-screen devices. In the remaining sections we focus on the example of result presentation and argue that the traditional approach of providing snippet text alongside results is inappropriate in a mobile context. Instead we propose a more effective solution based on the reuse of past queries as the basis for a more economic approach for *gisting* the meaning of results. We argue that this alternative is better suited to the display characteristics of mobile devices without compromising the informativeness of result snippets.

2. RELATED WORK

In the context of this paper, there are two separate strands of related work that are especially relevant when it comes to delivering information content to, or adapting information content for, mobile devices. The first strand concerns the general area of *web page adaptation* which involves automatically transforming or restructuring Web pages so they may be displayed more effectively on mobile devices. The second strand of related work is more focused on the presentation of search results on mobile devices and is obviously closely related to the work presented in this paper.

2.1 Web Page Adaptation for Mobile Devices

Many mobile search engines (including Google) seek to provide mobile users with access to normal Web content. But of course to do this, the content must be adapted so that it is compatible with mobile device displays. For example, the *Digestor* system [2] uses a *re-authoring* approach to transform a Web document using a range of design heuristics as well as text summarization, page categorization, the removal of irrelevant content and image reduction techniques. For example, Digestor's design heuristics tell it that keeping at least some images is important (usually the first and last image), that header tags (*H1-H6*) cannot be trusted as proper semantic headers for use as proxies for a block of text (instead text blocks are better summarised using the first sentence or phrase). Preliminary results suggest that while Digestor does a good job of preserving key content during re-authoring, the results are often not aesthetically pleasing when viewed on a small-screen device.

The *WEST* Browser (WEb browser for Small Terminals) [3], uses a technique called *flip zooming* which is a tile based $focus + context$ visualization technique for displaying web pages on handheld devices. A comparative evaluation carried out by the authors showed that the WEST browser provided users with a better overview and an easier search mechanism when compared to the HotJava browser. However, users also thought that the flip zooming interaction technique was quite difficult to use and took some time to get used to. The *PowerBrowser* system [4] provides a set of tools for searching, navigating, browsing, and input entry on small devices. Results of an evaluation carried out by the authors demonstrated that users experienced significant time-savings while using PowerBrowser for directed tasks.

In *WebThumb*, [18], the layout of pages is left untouched. Instead, graphical thumbnails are used to display whole pages and the browsing experience is improved by enhancing the normal interaction techniques available to the user. For example, a *pick up* tool enables users to extract elements from a page and display them in a separate window and *zooming* and *panning* tools allow users to take a closer look at content of interest.

The thumbnail concept is extended by [13]. The authors present a prototype application called *SearchMobil* which is able to partition a document into a number of different regions by examining the underlying structure of the page and provide a thumbnail overview of the document using these regions. These thumbnails are then annotated to show the location of query terms within the document with the aim of directing users to the most promising sections of the page. A user study carried out by the authors, showed that this approach was well-suited to fact-finding tasks. In [19], importance values are assigned to different segments of a web page in order to present mobile users with more compact search results and hopefully point users in the direction of more relevant results.

There are obvious challenges when it comes to adapting graphically-rich Web content for screen-poor mobile devices and in our opinion it is difficult to envisage an automated solution that will be capable of competing with the experience associated with mobile-specific content that has been designed for mobile devices.

2.2 Displaying Search Results on Mobile Devices

More directly relevant to the focus of this paper is the very specific challenge of how best to present search results for a mobile device. The standard Web strategy of presenting page titles, URLs and (hopefully) informative snippets of text is often adopted but comes with its problems, not the least of which is the high screen "real-estate" demands that snippet text imposes on mobile devices. One alternative that is also common place sees the elimination of snippet text altogether, leaving users at the mercy of often uninformative result titles. Clearly there is a need for a solution that can provide information regarding the relevance of specific results, but without consuming the screen resources of full snippet texts.

Work in this area has been rather limited to date. For example, [9] have looked at the general issue of the performance of Web searchers and their mobile counterparts on a range of different search tasks. The overall aim of the evaluation was to identify the impact of screen size on search performance and the results, unsurprisingly, pointed to a significant drop in search performance for mobile searchers. In addition the study also highlighted significant challenges for mobile searchers when it came to interpreting the usefulness of individual results according to their search needs.

This issue of how to help users to understand the value of recommended results is addressed by [10]. Instead of using standard snippet text approaches (which involve the extraction of a block of document text, usually related to the query) they use a set of key phrases, automatically extracted from result pages. The resulting key phrases provide for a more economic use of screen space and are at least as effective and informative as using long result titles.

3. THE STATE OF MOBILE SEARCH

Before continuing it is worth reflecting on the current state of the Mobile Internet and mobile search. And so in this section we will briefly review a number of important Mobile Internet developments over the past few years, developments that are responsible for increased levels of mobile search, in addition to the summary results from a recent review of some of the leading mobile search engines.

3.1 The Mobile Internet and Handsets

The late 1990's saw the advent of WAP (Wireless Application Protocol), the first generation of the Mobile Internet, promising users a new era of Mobile Internet Services. Unfortunately these promises rarely stood up to scrutiny and the content-light text-based services that were available did little to excite users about the potential of the Mobile Internet, especially when combined with low-bandwidth connections and expensive billing models.

In recent times however, the Mobile Internet has experienced something of a rebirth. Content has improved dramatically, offering users a wide range of rich-media services including colourful information pages, polyphonic ringtones and video on the go. In addition, the slow connections associated with the early Mobile Internet have been replaced by a much faster GPRS and 3G infrastructure providing users will almost instant access to content. Operators have also overhauled their billing practices to offer subscribers a much more cost-effective Mobile Internet service; instead of charging users for their time online, they are now only charged for the content they consume.

While all of these changes have led to significant improvements in the state of the Mobile Internet, perhaps the single most significant development has been the mobile handsets themselves. Gone are the tiny text-based, monochrome displays of the original WAP phones and in their place we have high-resolution, colour handsets with built-in browsing features and enhanced data input capabilities. In general, today, there are three types of mobile handset on the market: (1) standard WAP phones that offer high-resolution colour screens, albeit small screens, with predictive text input; (2) 3G smart phones with larger displays, enhanced browsing support (e.g. xHTML), and miniature keyboards; (3) PDAs with large colour displays, stylus/pen input, full HTML and Flash 6 support, and enhanced interaction features (e.g. full QWERTY keyboards).

3.2 Evaluating Mobile Search

The usage increases that have been concomitant with these improvements in handsets, bandwidth and content have led to an increase in mobile search, as users venture beyond operator portals to explore the growing content of the Mobile Internet. However, when it comes to helping users to locate information in the growing information space, improvements have been slow to come. For example, mobile search engines remain limited and fail to offer users a high level of user experience. To qualify these limitations, in the remainder of this section we outline the results of a recent evaluation of the state of mobile search.

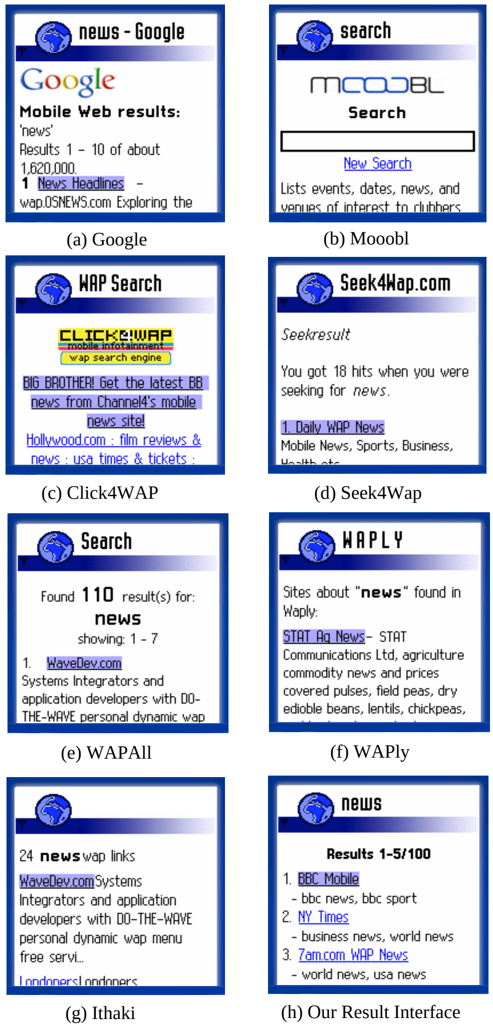
3.2.1 Methodology

To begin with, a representative sample of 7 mobile search engines were chosen as evaluation targets. Click4WAP, Google, Ithaki, Mooobl, Seek4Wap, WAPAll and WAPly were chosen for a number of reasons. First they represent a mixture of older search engines as well as the latest offerings; for example, Mooobl was launched as recently as June 2005. These search engines were also chosen because each focuses on the retrieval of mobile specific content, rather than attempting to retrofit standard HTML pages for mobile handsets. That said, it is worth noting that Google is distinguished by its indexing of xHTML pages as well as WML pages whereas the other engines primarily index WML pages. Furthermore, the Ithaki search engine is a meta search engine that combines results from WAPAll, FreoWAP, Indexcell, Click4WAP and WAPitOut.

A sample of 20 queries (e.g.: *cheap flights*, *tour de france*, *weather forecast*, *ringtones*) were submitted to each of the 7 search engines and their results were retrieved and recorded (See Appendix A for a full list of the 20 queries). Each result was manually assessed for its relevance to the target query and a range of relevance statistics were calculated for each engine including:

1. (NRR) The average number of results returned per query;
2. (1stRR) The average position of the 1^{st} relevant result in the result-lists for each query;
3. (%QwR) The percentage of queries for which the search engine return results;
4. (%QwRR) The percentage of queries for which a relevant result could be found in result-lists;

To illustrate the type of interfaces each of our sampled search engines present to their users we have included screen shots of the initial search results screen generated by each of the 7 search engines to the query *news* on a Nokia series 60 WAP phone; Figure 1 shows each of these screen shots. In addition to the above relevance considerations we will also assess the different ways in which these search engines present their result lists with a view to better understanding the presentation trade-offs they have come to adopt in light of the significant screen limitations of mobile devices, compared to their large-screen desktop and laptop relations.



(a) Google (b) Mooobl (c) Click4WAP (d) Seek4Wap (e) WAPAll (f) WAPly (g) Ithaki (h) Our Result Interface

Figure 1: Illustration of the Search Results Displayed on each of the 7 Search Engines in Response to the Query 'News'. Figure (h) in this Group Illustrates our Related Query Interface Approach.

3.2.2 Relevance and Coverage

The results corresponding to the above are presented in Table 1. The first thing to notice is the difference between Google and the other engines when it comes to the number of results retrieved per query on average. Google retrieves 126,000 results per query on average compared to a much lower average for the remaining engines; WAPly retrieves the least with 7 results per query, with Click4WAP the best of the rest at 38 results per query. These differences point to significant variations in the index size of the different engines, although it would be misleading to claim a simple direct correspondence between the average number of results retrieved per query and index size; we do not propose to consider this issue in detail here but the interested is referred to [12] for a related Web study. That said it is clear that Google's index size is significantly larger than the competing engines most likely owing to its coverage of xHTML content as well as WML content.

Google's indexing of xHTML pages as well as WML allows it to perform best of all. It has results to offer for all of the queries and returns relevant results for 85% of queries, with the top relevant result occurring at position 7 on average. However it is clear that the state of the pure WML search engines is somewhat less healthy, and given that a large number of mobile handsets are not yet equipped to handle xHTML content, this highlights an important problem for most Mobile Internet users. For example, when we look at the average position of the 1^{st} relevant result in result lists we see that Seek4Wap performs best with an average position of 1. However, we also see that while this search engine retrieves results for 90% of the queries it only offers a relevant result in 15% of queries. To put this another way, for the vast majority of queries (85%) Seek4Wap delivers an average of 10 irrelevant results (and no relevant ones), although when it does locate a relevant result it returns it in position 1. Compare this to Mooobl, which also retrieves a relevant result for 15% of the queries, but returns these relevant results at position 4 on average. However, since Mooobl only retrieves results for 20% of queries, it can at least claim to avoid returning lists of irrelevant results when relevant pages do not exist in its index. Overall Ithaki, with its meta-search strategy, performs best of the WML search engines, retrieving relevant results for 40% of queries (with an average position of 5 for the top relevant result) and retrieving irrelevant results for only 10% of queries.

Search Engine	RR	1stRR	%QwR	%QwRR
Google	126,000	7	100	85
Mooobl	21	4	20	15
Click4WAP	38	10	50	25
Seek4Wap	10	1	90	15
WAPAll	18	6	35	15
WAPly	7	7	15	10
Ithaki	9	5	50	40
$Mean_{WML}$	17.2	6	43.3	20

Table 1: General Properties of the Search Engine Results

In general, Google aside, the average performance of the 6 WML search engines is poor: they retrieve at least 1 relevant result for only 20% of queries and they return sets of irrelevant results for 23.3% of queries while providing no coverage for more than two-thirds of queries. Moreover, even when a relevant result is located for a query it is generally positioned low down in the result-list (position 17.2 on average) thus requiring mobile users to scroll through anything from 5 to 15 screens of results depending on screen-size and result presentation format.

3.2.3 Result Presentation

As well as examining whether the search engines return relevant results to our queries, we were also interested in how each search engine presents their results on the small-screen. If we look at general Web search, a user normally submits a query to a search engine and is presented with quite a large result-list, normally between 10 and 20 results per page. Each result usually consists of a result number/rank, a title, a short snippet of text from the result (usually contextualised for the query) and the result URL. An important implication of this concerns the large amount of screen space that normal search results demand and this type of format is unlikely to translate well onto mobile phones.

During our evaluation, in addition to the above relevance statistics, we also noted various presentation features for the different search engines. In particular we noted the presence of number/rank, title, snippet and URL information for each search result. The summary results are presented in Table 2 with 'Y' and 'N' indicating the presence or absence of these various features. The first thing to notice is that only Google's mobile search chooses to use all four presentation features for its results, with none of the WML specific engines following suit. Certainly Google's focus on xHTML content suggests a prioritisation of the larger screen smart phones and PDAs as its target market which seems to be supported by its use of rank, title, snippet and URL information in its result lists. Google's results are not well adapted to the smaller screen WAP phones that continue to dominate however, and arguably even modern smart phones will struggle to present more than 2 results per screen.

The WML-specific search engines adopt very different strategies with all of them making certain compromises when it comes to presenting each result. For example, at one extreme Mooobl only presents the title, where as at the other extreme WAPAll shows everything but the URL. In general all of the WML engines show title information, all but one drop rank information and none present the URL string. Interestingly two-thirds continue to present snippet information. This is surprising given the amount of space required to present even short snippet texts. The problem of course is that while dropping space hungry snippet text allows for a much more economic use of the mobile screen it does make it very difficult for the users to evaluate the relevance of a given result; searchers are left to rely on title text alone which is often not very informative.

Search Engine	Num/Rank	Title	Snippet	URL
Google	Y	Y	Y	Y
Mooobl	N	Y	N	N
Click4WAP	N	Y	N	N
Seek4Wap	N	Y	Y	N
WAPAll	Y	Y	Y	N
WAPly	N	Y	Y	N
Ithaki	N	Y	Y	N

Table 2: How Each Search Engine Presents the Search Results

To consider the screen-space economics of search results in more detail, in Table 3 we present information about the average length (in characters) of each result feature for the different search engines. As expected the average Google result is the most space hungry of the engines, consuming an average of 138 characters per result. In contrast, the majority of WML engines require less than half of this. Interestingly, two of the WML engines, Seek4Wap and WAPAll, use greatly truncated title text, requiring only 7 characters per title on average compared to an average of about 42 characters per title for the other engines. Overall Seek4WAP is the least space hungry of the engines tested. Its use of truncated title and truncated snippet text means that it requires an average of only 26 characters per result, which is even less that the average title text space needed by many of the competing engines.

In general then it should be clear that there is a major presentation issue when it comes to how best to present search result information in a way that is informative to users while at the same time sensitive to the screen limitations of most mobile phones. Considering that a standard Nokia series 40 mobile phone only fits approx. 6 lines or about 130/140 characters of text, while a Nokia series 60

Search Engine	Title	Snippet	URL	Whole Result
Google	17	62	31	138
Mooobl	59	-	-	59
Click4WAP	64	-	-	64
Seek4Wap	7	18	-	26
WAPAll	7	42	-	50
WAPly	19	65	-	84
Ithaki	50	-	-	50

Table 3: Average Length in Characters of each Search Result

mobile phone (smartphone) fits approx. 7 lines of text or approximately 200 characters of text per screen, most of the search engines examined will allow for only one or two results per screen.

3.3 Challenges for Mobile Search

The challenges for mobile search should be becoming clear. Existing search engines suffer from significant coverage and relevance issues with many queries either going unanswered or being answered by misleading result-lists containing irrelevant results. Coverage is likely to improve as search engines improve their ability to crawl the mobile Web. However, there are significant crawling issues to resolve going forward, issues that are quite different from those addressed in traditional Web search. For example, even though the mobile Web is many orders of magnitude smaller than the traditional Web, crawling remains a challenge because of the transient, short-lived nature of mobile content. Furthermore, mobile pages are much smaller than their Web counterparts and thus there is less information available as a source of indexing. Careful authorship is likely to be much more important than on the traditional Web if search engine indices are to be accurate. Finally, user queries are often vague in Web search (average query size tends to be between 2-3 terms [8]) but this is likely to be exacerbated by the limited input capabilities of mobile devices. So mobile queries are likely to be even shorter and more ambiguous that their Web counterparts. These issues probably explain, at least in part, the poor relevance reported in our preliminary evaluation above.

In addition to these coverage and relevance issues, presentation and interface design becomes much more critical in mobile search than in traditional Web search. Our study indicates that the interface priorities of most mobile search engines facilitate the display of only 1 or 2 results per screen on a typical handset, and those that can display more results do so by sacrificing important contextual information that a user will likely need to make a judgement regarding result relevance. Our focus in this paper is on the most space-intensive part of a typical result, its snippet text. Snippet text is a constant in modern Web search but, we believe it's an inappropriate luxury in mobile search. However we also accept that displaying just a title or a URL to represent each search result does not provide the user with enough meaningful information about the context of a given result. Hence the need for a more economic alternative for mobile search result presentation.

4. INTELLIGENT RESULT GISTING

The objective of this work is to present and evaluate an alternative approach to search result gisting that enjoys the informativeness of snippet text while providing for a more economic use of limited screen real-estate. The core idea behind this approach is to replace result snippets with a much shorter text representation that is made up of the terms of related queries that have led to the selection of a particular result in the past. This approach has been made possible as a direct consequence of a community-based personalized meta-search engine called I-SPY, which records the queries and search results of different communities of users.

4.1 I-SPY & Collaborative Web Search

I-SPY is a community-based meta-search engine that provides its users with search results that are informed by the past search behaviour of a community of like-minded users. In essence I-SPY selectively re-ranks search results according to the learned preferences of a community of users, promoting those results that are likely to be relevant to the current query using a record of search behaviors carried out by the community. Specifically, I-SPY monitors users selections or hits and maintains a record of queries and result selections [15, 16]. Each time a user selects a result, p, for a query q, I-SPY updates a community profile to reflect this new selection. This community profile forms the basis of I-SPY's relevance metric. It records all of the queries submitted and results selected by a particular community and the relevance of a page p to a query q for a community c is calculated as the probability that page p will be selected for query q. This probability is estimated as the proportion of times that p has been selected for q in c in the past. Further details can be found in [15].

I-SPY maintains a separate profile for different communities of users; e.g., searches that originate on a motoring web site are kept separate from searches that originate from a wildlife web site. This separation of communities, coupled with I-SPY's approach to relevance, allows I-SPY to predict that users of the motoring web site are more likely to be looking for sports car sites when they enter the query *jaguar*, whilst users of the wildlife site are most likely to be looking for information on large cats for the same query, for example. In this way I-SPY can disambiguate effectively between vague queries and a range of user trials have shown how I-SPY can generate superior result rankings than leading search engines such as Google. For example, a recent study has shown how I-SPY, working with Google as its underlying search engine, can reduce the percentage of search failures and improve the positioning of relevant results when compared to Google [14].

4.2 Query Reuse for Result Gisting

One of I-SPY's distinguishing features is its storage of past search session information such as the queries that have been submitted and the results that have been selected for these queries. The work of [1] proposed how this information could be leveraged for a novel form of query recommendation in which related queries are recommended alongside certain search results that have previously been selected. For each search result p_k, selected for some target query q_T, I-SPY can generate a set of related queries, $q_1, ...q_n$, which have previously led to the selection of p_k. Our idea for result gisting is based on a modification of this query recommendation technique. Specifically, we combine the terms used in these queries to provide a compact yet meaningful summary of the corresponding search result.

Figure 2 illustrates this idea in action. If we take, for example, the query *Java* we can see that the first search result, *Java Sun Technology* is associated with queries such as *j2sdk1.5* and *java tutorials*. These 'related queries' help to inform the user about other contexts in which this result was selected and at the very least tell the user that the result was found to be relevant for users looking for the latest SDK and java tutorials.

In previous preliminary evaluations [5, 6], we have considered whether these related queries can be used as an alternative to snippet-text to gist search results. The empirical evidence from both of these previous evaluations demonstrated that the related queries were as informative as snippet text and offered the potential for

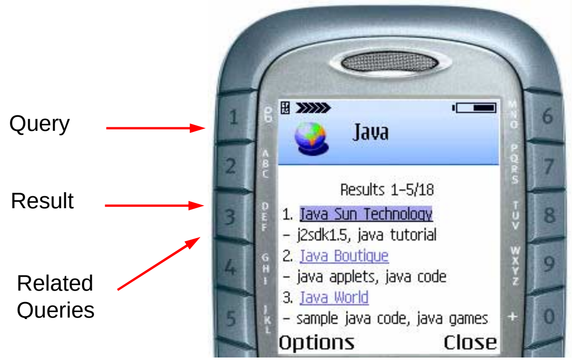


Figure 2: Illustration of the Results and Related Queries Generated for the Query 'Java' on a Mobile Phone

a significant space saving. However these evaluations were limited because they did not involve the judgments of real users in realistic search scenarios. Thus in this section we describe the results of a new live-user evaluation that focused on how informative users perceived these related queries to be as an alternative to search result snippets. But first, we will briefly review the previous offline results presented in [6] to provide a context for our latest study.

4.3 An Offline Evaluation

We evaluated the usefulness of related queries as an alterative to snippet-text for result-gisting using data from I-SPY search logs. The data for the evaluation consisted of 684 result-pages that were selected by searchers in response to more than 2,600 queries. Each of these pages had at least 2 related queries associated with it as well as a unit of snippet text (generated by Google). The primary goal of our offline evaluation was to determine how well these related queries represented the page in question, relative to its snippet text. To do this we supposed that the representativeness of a set of terms relative to some page could be measured by the position of the page in the result-list generated by some search engine when using these terms as a query. Hence, in our evaluation, for each target page p we transformed its set of related queries and piece of snippet text into two new search engine queries; one based on the related query terms and one based on an equivalent number of snippet text terms. We then submitted the queries to HotBot and compared each query according to the rank of the target page in the corresponding result-list.

4.3.1 *Query Generation*

The most crucial part of the evaluation was the generation of the test queries. In all, six query generation strategies were tested: two that produced queries from the terms contained in the related queries for a page and four that used terms from the page's snippet text. Strategy $RQ1$ produces a test query by concatenating the related query terms into a single query. Strategy $RQ2$ uses a similar approach but duplicate terms are removed.

The snippet text conversion was slightly more complicated. To generate the test queries we parsed the snippet text to remove stop-words and special characters and then selected terms from the remaining snippet text using four different strategies. In strategy $S1$ we select a random set of k terms, where k is the number of terms in the test query produced by $RQ1$. Strategy $S2$ selects the top k most common terms in the snippet text, where k is the number of terms in the test query produced by $RQ2$. For strategy $S3$ we again select a random set of k terms but this time k is the number of terms produced by $RQ2$. Finally, strategy $S4$ selects the top k most common terms in the snippet text, where k is the number of terms produced by $RQ1$.

4.3.2 *Relevance Assessment*

After submitting each test query to HotBot, we examined the top 500 HotBot results only and compared the position of p, the target result-page, in the result-lists produced for each test query. The higher p is in the result-list the more representative the test query must be as an indicator of $p's$ content and hence the more representative the related queries or snippet text. Along with this positional information we also examined the percentage of results matched by each test query strategy as well as the average length in number of terms of the related queries vs. snippet-based strategies.

4.3.3 *Results*

First we looked at the average position of each target page p in the result-lists produced by HotBot. Table 4 shows that both related query strategies, $RQ1$ and $RQ2$, perform very well. $RQ2$ locates p at an average position of 39 in the result-list compared to the best performing snippet-based strategy which locates p at an average position of 103. Note that if the target result cannot be found, p is given a default position value of 501 (because we are only examining the first 500 HotBot results).

Test Query	RQ1	RQ2	S1	S2	S3	S4
Avg. Position	46	39	144	118	180	103

Table 4: Average Position of p in Result-Lists

The poor performance of the snippet text strategies could have been due to frequent penalties being incurred when p was not present in the top 500 results. To understand this we examined the percentage of results found for each test query strategy in the top 500 HotBot results, see Table 5. The related query strategies succeed in producing result-lists that contain p for between 92% and 94% of queries, a significant improvement when compared to the snippet text strategies which return p for between 66% and 80% of the test queries.

Test Query	RQ1	RQ2	S1	S2	S3	S4
Percentage (%)	92	94	73	78	66	80

Table 5: Percentage Found in Top 500 HotBot Results

These results suggest that the terms contained in the related queries are more representative of the pages they refer to than an equivalent number of terms taken from the snippet text associated with these pages. They also suggest that by using related queries instead of snippet text we can achieve a significant saving in display-space. For example, as shown in Table 6, our related queries contained only 4 unique terms on average, compared to snippet texts with an average of 35 terms or 21 terms with stop-words removed.

Test Query	RQ1	RQ2	Snippet	Parsed Snippet
Avg. Num Terms	6	4	35	21

Table 6: Average Number of Terms in Test Queries

The results of the above evaluation suggest the use of related queries as an economical alternative to snippet text for result gisting. The terms contained within related queries appear to have the potential to better capture the essence of their associated pages than the terms in the snippet texts, and so may serve to be a more informative gisting approach. These *good* terms have the advantage that they were used in situations where a given page was ultimately selected ,the same is not true for snippet text terms. Moreover related query terms have been generated (in I-SPY) by a community of like-minded searchers, which should help to constrain the possible interpretations of possible query terms. Moreover, related query terms take up a small fraction of the screen space associated with the display of snippet text, which is a major advantage for the provision of mobile search.

Of course the limitation of the above study is obvious. It does not involve live-users during relevancy assessment and so it is not clear as to whether real users would likely behave in a similar way. In the following section we describe our live user study designed to make some progress in this regard. In particular it is designed to determine if real users accepted our related query interface as an effective interface for displaying search results on a mobile phone.

4.4 A Live User Evaluation

In this new study we asked users to evaluate three different interfaces for displaying search results on a mobile phone. The goal of the evaluation was to understand how users judged the informativeness space trade-off between snippet text and our related queries approach to result gisting. In brief, the evaluation asked a set of test users to evaluate three different result-list interfaces for a particular set of result pages and their related queries and snippet texts.

4.5 Phase 1 - Related Query Generation

The first part of this evaluation involved the generation of 3 different result-lists, for a set of 18 result pages, but each focusing on a different presentation style. The first style presented result title information only (Interface 1). The second presented title plus snippet text (Interface 2). And the third, used our new approach to gisting, presenting title plus a related query string (Interface 3).

For the third presentation style we needed a source of realistic related queries as the basis of our query string. To generate these related queries we asked 5 users to view a set of result pages and formulate a set of queries they would enter in a search engine if they were looking for the page in question. Each participant examined 18 pages relating to 6 different AI & Computer Science conferences including WWW, SIGIR and IJCAI; that is we presented 3 different web pages per conference. We asked the participants to open and view each web page and devise 5 different queries they would enter in a search engine if they were looking for the web page in question. These queries constituted the *related query database* and in this evaluation scenario corresponded to the sort of queries that a community of computer science students might enter when looking for conference information.

To generate the related query string for each result page, we extracted the top 3 most frequent queries entered for that page and then computed the union of these terms to produce a single related query string. For example, the top three terms entered by the participants for the *IJCAI 2005 General Conference* web page were, *IJCAI 05*, *IJCAI Edinburgh* and *IJCAI 2005 Info*. The combined query string was *IJCAI 05 IJCAI Edinburgh IJCAI 2005 Info* and the related query string we were left with after removing duplicate terms was *IJCAI 05 Edinburgh 2005 Info*.

4.6 Phase 2 - Interface Evaluation

During this part of the evaluation we presented our three different styles of result-list to 120 users from a computer science / IT background.

4.6.1 Methodology

Each participant was presented with 6 sets of our 3 different mobile phone interfaces with each of the 6 sets presenting the results of a different search for a particular computer science / AI conference; each set of results was presented using an interface designed for a Nokia series 60 WAP phone. And within each set we presented the result listing using *title* information only, *title and snippet* information, and *title and related query string*. The titles we used were extracted directly from each of the 18 result pages described earlier and the snippet text we used was generated by Google for each of the 18 result pages.

Each user was then asked to answer two questions for each of the 6 sets of results bearing in mind that as a searcher they might be interested in different types of information that was not declared as part of their original query; for example, the query *WWW 2005* might be used by a searcher looking for the *call for papers* or *registration information*. The two questions presented to users were: (1) Which interface provides the most useful information about the search results returned? (2) Which style of presentation strikes the best balance between use of screen space and information conveyed?

Figures 3, 4 and 5 show some of the different interface sets presented to users of the study. The curly bracket to the left of the interface sets illustrates the approximate screen size of a typical xHTML-enabled mobile phone. We included this bracket to help users visualize how much information actually fits within a single screen on the mobile device. By examining the information within each of the different interface sets you can see that the information displayed to users is of varying quality. For example, figure 3 shows the 3 different interfaces for the WWW 2005 conference pages. In this case, each of the titles, snippet text and related queries are quite informative and easily distinguishable. However if we look at figure 4, IJCAI 2005, we see that all the title text is the same and in figure 5, ICCBR 2005, all of the snippet text is the same even though the results point to different pages. This is a common problem on both the World Wide Web and the Mobile Internet and is primarily caused by improper web authoring. It is especially important in these cases to provide some additional contextual information to help users understand and distinguish between relevant and irrelevant search results. Related queries can provide this additional context.

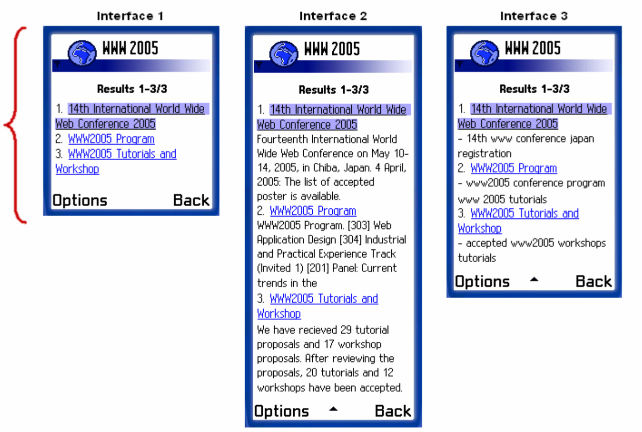


Figure 3: Illustration of the 3 different interfaces shown to users of the study for the WWW 2005 conference

In total our questionnaire resulted in 112 completed users sessions. Each completed session involved the user selecting one in-

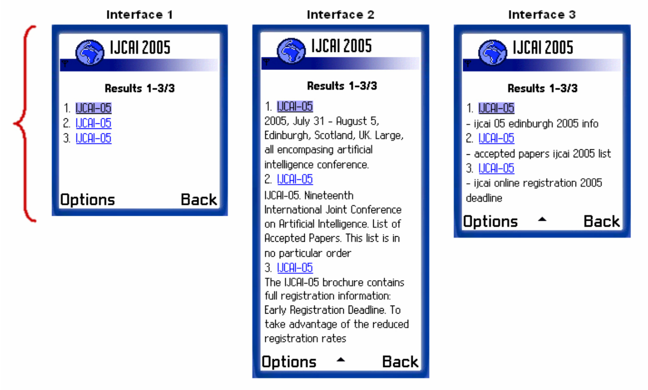


Figure 4: Illustration of the 3 different interfaces shown to users of the study for the IJCAI 2005 conference

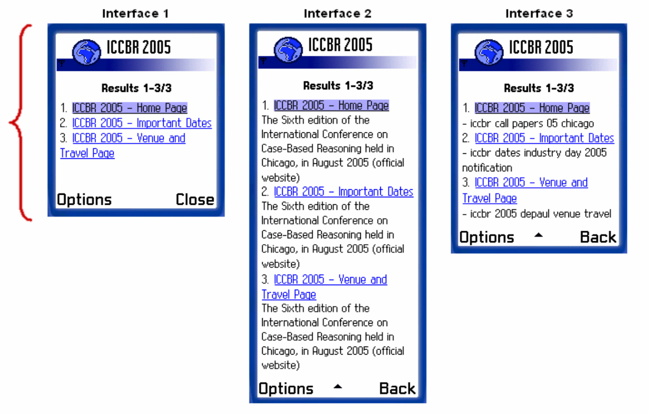


Figure 5: Illustration of the 3 different interfaces shown to users of the study for the ICCBR 2005 conference

terface in answer to each of the above questions for the 6 sets of result-lists; thus, each session produced 12 user interface selections. These formed the basis of the results presented in the following section.

4.6.2 Results

Figure 6 shows the overall performance of each of the three interfaces. The *overall performance* of each interface refers to the percentage of user selections each interface received, averaged across both questions. We can see from the graph that interface 3, the related query interface, performs best overall, with 54% of the selections received for this interface across both questions. Interface 2 received 34% of the selections, while interface 1 performs worst with just 11% of user selections. This suggests that interface 3 might provide a suitable interface type in mobile search environments, striking a good balance between use of screen space and the quality of information displayed. Interface 1 performs worst overall thus confirming our prediction that providing just the title of each search result does not provide the user with enough information about the context of the result in question.

Figure 7 illustrates the performance of each of the interfaces for each of the two questions asked. On the issue of which interface provide the most useful information (question 1) there is a clear preference for interface 2 (title and snippet text), which received 60% of user votes. Interface 3 (title and related queries)

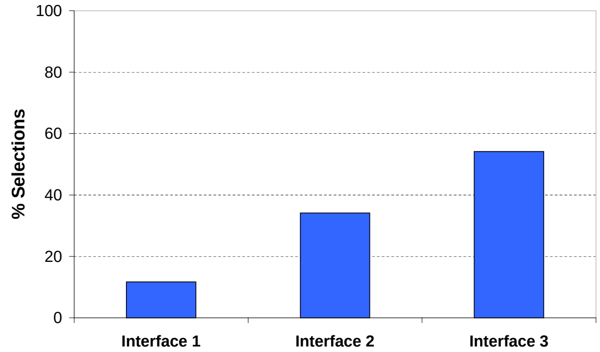


Figure 6: Illustration of the Overall Performance of each of the Three Interfaces Across Both Questions

attracted only 33% of the votes and interface 1 only 7% of the votes. This is not surprising and the results correlate precisely with the quantity of information provided with each search result (see Table 7). The content-rich snippets of text used by interface 2 certainly provide more information to the user about the result than the more economical related query strings used by interface 3. However, this evaluation question purposefully ignores the issue of the informativeness-space trade-off.

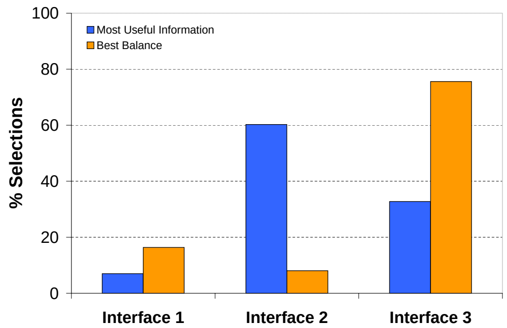


Figure 7: Illustration of the Performance of each of the Three Interfaces for each of the Individual Questions Asked

Interface	Presentation Style	Result Length
1	Title Only	28
2	Title + Snippet Text	167
3	Title + Related Queries	73

Table 7: Average Length of the Search Results in Characters

Question 2 focuses on this trade-off explicitly by asking users to evaluate which interface presents the best balance between information and space usage. And on this issue there is a strong preference for interface 3 (title and related query strings), which attracted over 75% of user preferences. In fact on this issue the traditional result-list presentation approach epitomised by interface 2 (title and snippet text) performs worst of all, attracting only 8% of user preferences. Indeed we find that interface 1 (title only) performs better, attracting 16% of votes, which is consistent with a preference (over title plus snippet text) for this style of result presentation format among the current set of WML search engines.

We carried out a two-way analysis of variance (ANOVA) to determine whether our results were statistically significant. The design was a 2 x 3 one with question (question 1 or question 2) and interface type (interface 1, interface 2 or interface 3) being the within-subject variables. The ANOVA test revealed a significant main effect of interface type $[F(2, 666) = 169.5, p < 0.001]$, and a highly significant interaction between the question and interface type $[F(2, 666) = 218.2, p < 0.001]$. We carried out Tukey's post-hoc comparisons to determine if the interaction effects found were reliable, the results of which showed that there were reliable differences between all of the questions and all of the interfaces.

In summary, these results suggest that providing title information alone is not optimal; titles on their own lack sufficient detail to be truly informative for the average searcher. At the same time, title plus snippet information, while much more informative is not appropriate for mobile handsets because of its high space demands. The evidence points to interface 3, which combines title information with our novel related query strings, as providing a better balance between information and space usage that is suitable to unique characteristics of mobile devices; interface 3 requires less than half of the screen space of interface 2, for example (see Table 7).

5. CONCLUSIONS & FUTURE WORK

As the Mobile Internet continues to develop at a pace, mobile search is likely to become a more important way for users to access information, just as Web search is the primary mode of information access on the Web today. However, the limitations of modern mobile handsets introduces a number of crucial challenges when it comes to delivering useful and usable search engine services. For example, one of the main issues concerns the manner in which search results are displayed. In this paper we have argued that traditional presentation styles are not optimal through an extensive study of 7 existing mobile search engines. We have proposed using related queries as a more economical alternative to the use of snippet text for displaying search results and as a more informative alternative to displaying result titles alone. We have included the results of two separate evaluations including one live-user trial. These suggest two important conclusions: first, they indicate that related queries do provide an informative alternative to snippet text; second they also suggest that users judge the use of related queries to provide a better balance between informativeness and screen-space on mobile handsets.

The findings of our mobile search engine study and the results of our related query evaluations point to some very interesting avenues for future research. At present, we are pursuing a number of different areas relating to both mobile search and the Mobile Web in general. Regarding our work in the area of search result presentation on mobile devices, our evaluations to date have yielded very positive results but we are aware that our studies are limited in some regards. This is a work in progress and we understand that more quantitative evaluations are needed in order for us to obtain a more objective evaluation. Therefore, as our next step, we plan to carry out a quantitative user evaluation where we ask users to perform real search tasks in a more dynamic and interactive setting.

We also plan to explore our related query technique in more detail. In particular we want to add more intelligence into both the generation and display of the related query strings. We also want to investigate the use of other non-textual interface types for mobile search result presentation. Finally, in a separate but related area, we are exploring the *dynamics* of the Mobile Web, in particular the evolution of content in the mobile space. Early results indicate that the Mobile Web is highly dynamic, presenting a whole new set of challenges for the Mobile Internet community.

6. ACKNOWLEDGEMENTS

The authors would like to thank all of the volunteers who took part in our user evaluation. This material is based on works supported by the Science Foundation Ireland under Grant No. 03/IN.3/I361.

7. REFERENCES

[1] E. Balfe and B. Smyth. Improving Web Search through Collaborative Query Recommendation. In *Proceedings of 16th European Conference on Artificial Intelligence*, pages 268–272, Valencia, Spain, 2004.

[2] T. W. Bickmore and B. N. Schilit. Digestor: Device-Independent Access to the World Wide Web. *Computer Networks and ISDN Systems*, 29(8-13):1075–1082, 1997.

[3] S. Björk, L. E. Holmquist, J. Redström, I. Bretan, R. Danielsson, J. Karlgren, and K. Franzén. WEST: A Web Browser for Small Terminals. In *UIST '99: Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, pages 187–196. ACM Press, 1999.

[4] O. Buyukkokten, H. Garcia-Molina, A. Paepcke, and T. Winograd. Power Browser: Efficient Web Browsing for PDAs. In *CHI '00: Proceedings of the Conference on Human Factors in Computing Systems*, pages 430–437, 2000.

[5] K. Church, M. T. Keane, and B. Smyth. An Evaluation of Gisting in Mobile Search. In *ECIR '05: Proceedings of the 27th European Conference on Information Retrieval*, pages 546–548, Santiago de Compostela, Spain, 2005.

[6] K. Church, M. T. Keane, and B. Smyth. Towards More Intelligent Mobile Search. In *IJCAI '05: Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 1675–1676, Edinburgh, Scotland, 2005.

[7] N. Gohring. Mobile phone sales topped 800 million in 2005, says IDC. http://wireless.itworld.com/4267/060127mobsales/, January 2006. Last Checked 28th April 2006.

[8] B. J. Jansen, A. Spink, and T. Saracevic. 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, 2000.

[9] M. Jones, G. Buchanan, and H. W. Thimbleby. Sorting out Searching on Small Screen Devices. In *Proceedings of the 4th International Symposium on Human Computer Interaction with Mobile Devices*, pages 81–94, 2002.

[10] S. Jones, M. Jones, and S. Deo. Using Keyphrases as Search Result Surrogates on Small Screen Devices. *International Journal of Personal and Ubiquitous Computing*, 8(1):55–68, 2004.

[11] P. Kendall. Worldwide Cellular User Forecasts, 2005-2010. Wireless Network Strategies, Strategic Analytics, 2006.

[12] S. Lawrence and C. L. Giles. Searching the World Wide Web. *Science*, 280(5360):98–100, 1998.

[13] N. Milic-Frayling, R. Sommerer, K. Rodden, and A. F. Blackwell. SearchMobil: Web Viewing and Search for Mobile Devices. In *WWW (Posters)*, 2003.

[14] B. Smyth, E. Balfe, O. Boydell, K. Bradley, P. Briggs, M. Coyle, and J. Freyne. A Live-User Evaluation of Collaborative Web Search. In *IJCAI '05: Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 1419–1424, Edinburgh, Scotland, 2005.

[15] B. Smyth, E. Balfe, J. Freyne, P. Briggs, M. Coyle, and O. Boydell. Exploiting Query Repetition & Regularity in an

Adaptive Community-Based Web Search Engine. *User Modeling and User Adapted Interaction*, 14(5):383–423, 2005.

[16] B. Smyth, J. Freyne, M. Coyle, P. Briggs, and E. Balfe. I-SPY - Anonymous, Community-Based Personalization by Collaborative Meta-Search. In *Proceedings of the 23rd SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 2003.

[17] TechNews. Mobile Web Browsing May Become Dominant Internet Platform. http://www.technologynewsdaily.com/node/2501, April 2006. Last Checked 28th April 2006.

[18] J. O. Wobbrock, J. Forlizzi, S. E. Hudson, and B. A. Myers. WebThumb: Interaction Techniques for Small-Screen Browsers. In *UIST '02: Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*, pages 205–208. ACM Press, 2002.

[19] X. Xie, G. Miao, R. Song, J.-R. Wen, and W.-Y. Ma. Efficient Browsing of Web Search Results on Mobile Devices Based on Block Importance Model. In *PerCom 2005: Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications*, pages 17–26. IEEE Computer Society, 2005.

APPENDIX

A. LIST OF THE 20 QUERIES SUBMITTED TO THE MOBILE SEARCH ENGINES

1. Jobs Ireland
2. Cheap Flights
3. Boston
4. Learn Italian
5. Premiership Results
6. Tour de France
7. BBC TV Listings
8. News
9. Weather Forecast
10. GI Diet
11. Britney Spears
12. Buy iPod
13. Music Downloads
14. Harry Potter
15. Java Tutorial
16. David Beckham
17. Big Brother
18. Banking 365
19. Recipes
20. Ringtones