Mobile Information Access: A Study of Emerging Search Behavior on the Mobile Internet

KAREN CHURCH and BARRY SMYTH
University College Dublin
and
PAUL COTTER and KEITH BRADLEY
ChangingWorlds

It is likely that mobile phones will soon come to rival more traditional devices as the primary platform for information access. Consequently, it is important to understand the emerging information access behavior of mobile Internet (MI) users especially in relation to their use of mobile handsets for information browsing and query-based search. In this article, we describe the results of a recent analysis of the MI habits of more than 600,000 European MI users, with a particular emphasis on the emerging interest in mobile search. We consider a range of factors including whether there are key differences between browsing and search behavior on the MI compared to the Web. We highlight how browsing continues to dominate mobile information access, but go on to show how search is becoming an increasingly popular information access alternative especially in relation to certain types of mobile handsets and information needs. Moreover, we show that sessions involving search tend to be longer and more data-rich than those that do not involve search. We also look at the type of queries used during mobile search and the way that these queries tend to be modified during the course of a mobile search session. Finally we examine the overlap among mobile search queries and the different topics mobile users are interested in.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search process, query formulation*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services

General Terms: Experimentation, Human Factors, Measurement

Additional Key Words and Phrases: Mobile browsing, mobile Internet, mobile search, log analysis

This material is based on works supported by the Science Foundation Ireland under Grant No. 03/IN.3/I1361.
Author's address: K. Church, University College Dublin, Dublin Ireland; email: karen.church@ucd.ie.

ACM Reference Format:
Church, K., Smyth, B., Cotter, P., and Bradley, K. 2007. Mobile information access: A study of emerging search behavior on the mobile Internet. ACM Trans. Web, 1, 1, Article 4 (May 2007), 38 pages. DOI = 10.1145/1232722.1232726 http://doi.acm.org/10.1145/1232722.1232726

1. INTRODUCTION

Mobile phones are now poised to overtake more traditional information access devices (e.g., desktop and laptop PCs) as the dominant platform for Internet information access in many markets. Recent estimates from Wireless Intelligence put the total number of cellular subscribers at 2.5 billion in September 2006 and predict that by 2007 there will be approximately 3 billion subscribers [Cellular News 2006]. Indeed, at the time of writing, a new report by Ipos Insight [2006] highlights how mobile phone ownership is now reaching saturation levels in many areas of the world. In East Asia, for example, more than 90% of all households now own at least one mobile phone. In Western Europe, the figure is 80%, while in the US, 3 out of every 4 households own a mobile phone. Indeed, globally, more than two-thirds of the households that own a mobile phone actually own multiple handsets with the average number of handsets in these households currently at 2.2. At the same time an increasing number of mobile subscribers are using their mobile phones to access the mobile Internet. According to the Ipsos Insight study, globally, 28% of mobile phone subscribers have used their phones to browse the Internet (just over a 3% increase from 2004 figures). In Japan, 40% of subscribers use the mobile Internet, while in Europe, France and the UK are leading the way; according to a 2005 Forrester Research study, 21% of European mobile subscribers use mobile Internet services at least once per month (see e.g., van Veen et al. [2005]). Moreover, this overall pattern of growth is driven by users in the 35+ age-group, indicating that mobile Internet usage is no longer dominated by the early-adopter subscriber segment. Growth in MI usage in this higher ARPU segment also suggests significant revenue growth opportunities going forward for mobile operators.

Given this growth in mobile Internet usage and mobile information access in general, it is likely that the mobile Internet will rapidly become a vital source of anytime, anywhere access to information for hundreds of millions of users. How will these users locate the information they are looking for? Generally speaking, there are two basic approaches to locating information: users may browse for information or they may use queries to search for information. During the early years of the Internet, directory services like Yahoo! meant that browsing served as the dominant form of information access. A similar effect is seen on the mobile Internet today with the dominance of operator-owned portals leading most subscribers to browse for content in order to satisfy their information needs. However, today search has come to dominate on the traditional Web and, in this article, we will argue that recent evidence points to the increasing role that search is playing on the mobile Internet.

To frame our research, it is important to understand the information access habits of mobile users and the particular challenges that they may face when it comes to accessing mobile information and content. Previous research, for example, has highlighted some of the problems inherent in mobile information browsing (e.g., the so-called *click-distance problem* discussed by Smyth [2002] highlighted the long navigation times faced by most mobile portal users) and their resulting solutions (such as mobile portal personalization is now a widely adopted solution that is capable of reducing portal click-distance by adapting generic portals to the needs of individual subscribers based on their browsing habits [Smyth and Cotter 2002a, 2002b]). An increased reliance on mobile search will introduce a new set of somewhat related challenges. While many of these challenges may have their origins in Web search, it is likely that the characteristics of the mobile Internet (e.g., device input and output limitations) will result in these problems being exacerbated and their effects more acutely felt by mobile subscribers.

In this article, we describe some of the results of a comprehensive survey of mobile Internet usage, drawing on the usage data generated by more than 600,000 European mobile Internet users towards the end of 2005. We are especially interested in their mobile information access habits, and, in this article, we highlight some important emerging trends in mobile search. Recent forays into the mobile Internet by the Web search giants (e.g., Google, Yahoo, AOL) have seen attempts to retrofit existing Web search technologies for the mobile Internet often with just a simple change of presentation format to accommodate the small screens available on mobile devices. While this may well be the most expeditious approach to delivering mobile search, we doubt whether it is the right way forward; similar strategies have led to significant user experience problems in the past as operators launched their original WAP portals. To address device issues and exploit the usage patterns of mobile searchers, we believe that the ultimate solution to mobile search is likely to be very different from current Web search-inspired solutions.

At present very little is understood about the nature of mobile search and the needs of mobile searchers. Our study seeks to learn more about the mobile search activities that are emerging among the usage patterns of mobile subscribers. We rely on more than 400,000 query-based searches for more than 30 different mobile search engines to develop a variety of statistics, from the devices that are used to search for mobile information to the type of queries that users tend to submit and the characteristics of their sessions. We also compare this search behavior to mobile browsing behavior and, moreover, we consider the striking parallels between current mobile search and early Web search. We go on to compare these early stages of mobile search with Web search behavior from the late 1990's with reference to a number of historic Web search studies. We view this analysis as an important first step on the road to better understanding mobile search, a journey that we believe will lead us to identify more innovative approaches to mobile search, approaches that are better adapted to the needs of mobile searchers and the characteristics of the mobile Internet.

2. RELATED STUDIES

In this article, we will focus on mobile information access in general, looking at both browsing and search behaviors on the mobile Internet. A significant part of this will include a detailed study of emerging search behavior on the mobile Internet, and, in this context, we will begin by reviewing a growing body of literature on the analysis of Web search which has developed over the past ten years. This will help somewhat to clarify many of the consistent behavioral characteristics that appear to underpin Web search as well as to demonstrate some of the changes that have been evident as Web search has become mainstream. Overall these studies share a similar basic methodology.

2.1 The Excite Studies 1997–2001

Perhaps the best known Web search studies have been conducted by Spink et al. [2001, 2002], Jansen et al. [2000], and Wolfram et al. [2001] on query logs from the Excite search engine. The seminal study, carried out on data generated for one day in March 1997 [Jansen et al. 2000], highlighted many of the now well accepted peculiarities of Web search that set it apart from traditional information retrieval systems: vague queries and brief search sessions, the minimal use of advanced search features and query modification operators, skewed search interests, etc. For example, the March 1997 study [Jansen et al. 2000] involved an analysis of 51,473 queries from 18,113 users, and the authors reported an average query length of only 2.21 terms with $< 4\%$ of queries containing 6 terms or more. Searchers rarely availed themselves of more advanced search features such as boolean operators or other forms of query modifiers, with less than 18% of queries including such features. In turn, search sessions were short, with users submitting 2.8 queries per session and looking at only 2.35 result pages per query; in fact more than 50% of searchers did not access results beyond the first page. The authors also found a skewed term distribution with a few terms used very often and a long tail of rarely used terms; for example, out of a total of 21,862 unique terms, 9,790 terms occurred just once. Finally, this initial analysis of Web search found searches for adult-related content to be prominent with 25% of the top 63 subject terms categorised as adult-related.

A second Excite study from September 1997 adopted a similar methodology, this time involving 1,025,910 queries posed by 211,063 users (see Spink et al. [2001]. The results are consistent with the earlier study in some respects, for example, an average of 2.16 terms per query, but show some behavior changes in other regards. For example, there was an increase in the number of users (71%) that ventured beyond the first page of results. There was also a further decrease in the use of advanced search features with less than 5% of queries involving boolean operators. The authors also reported an increasing interest in entertainment-related information and a slight move away from adult-related material.

The third Excite study [Wolfram et al. 2001], based on analysis of more than 1 million queries by over 200,000 users from 1999, showed little deviation from the core 1997 statistics. Queries continued to have approximately 2 terms on average, and the use of advanced search features remained rare (8% of

queries). However, an increase in the percentage of search sessions involving just a single query was reported (up from 48% of sessions in the 1997 studies to 60%), and 43% of users chose not to venture beyond the first page of results, which was more in line with the March 1997 study (approximately 50%) than the September 1997 study (29%). The 1999 study also reported an increase in search activity related to e-commerce, travel, and employment information and away from adult-related interests.

A final Excite study [Spink et al. 2002] (1,025,910 queries across 262,025 search sessions by over 200,000 users) from April 2001 showed little change except for a slight increase in the number of terms per query (2.6) and a slight increase in the percentage of users viewing just a single page of results (50%). This 2001 study further highlighted the shift in user interests away from adult content and towards e-commerce.

2.2 Altavista 1998 and 2002

By 1998, Altavista was the leading Web search engine and a study by Silverstein et al. [1999], based on 43 days worth of query logs during August and September of that year, led to a wide range of results that were broadly in line with the Excite studies referred to previously: 2.35 terms per query on average; a strong bias towards adult-related search with 50% of the top 25 queries linked to adult content; less than 15% of searches caused users to venture beyond the first page of results; and advanced search features were used in only 20% of searches. A follow-up study based on an analysis of 1,073,388 queries from 369,350 search sessions collected during a 24-hour period in September 2002 highlighted a slight increase in average query length for Altavista searchers (up to 2.92 terms) and a significant increase in the frequency with which users would look beyond the first page of results (27% up from 15% in 1998) [Jansen et al. 2005]. Interestingly, in the 1998 study, 77% of search sessions were made up of a single query. By 2002, more than 50% of sessions contained 2 or more queries with users modifying their original queries 52% of the time (compared with 20% of the time in 1998).

2.3 European Web Search—Fireball and AllTheWeb

Two largely European Web search engines namely Fireball (www.fireball.de), a German search engine, and AllTheWeb, a highly-rated European search engine (www.alltheweb.com), provided an opportunity to investigate the search patterns of a mainly European community of Web searchers. The Fireball study used 31 days of data from 1–31, January 1998, which covered more than 16 million queries and 3 million Web sites [Höscher and Strube 2000]. While results indicated an even shorter average query length by Fireball searchers than their Excite and Altavista counterparts (1.66 terms per query), other statistics on the low use of advanced search features (more than 97% of Fireball queries used no advanced search operators) and the reluctance of Fireball searchers to look beyond the first page of results (just over 59% of searchers focused their attention on the first page of results only) remained largely in line with these other studies.

The AllTheWeb study [Jansen and Spink 2005] was based on two sets of data from February 2001 and May 2002. Each dataset was based on 24 hours of usage and covered approximately 200,000 searchers and 1 million queries. This study highlighted a decline in average query length from 2001 to 2002 with the percentage of single-term queries, for example, growing from 25% in 2001 to 33% in the 2002 dataset. Search sessions also reduced in duration, and there was a notable increase in the variety of search topics with an attendant decline in the level of activity related to adult searches.

2.4 Towards an Analysis of Mobile Search

At this point, it is worth summarizing some of the main findings to come out of these early Web search studies.

(1) Search queries tend to be short and thus have the potential to be vague or ambiguous; most studies report average query lengths of between 2 and 3 terms.
(2) Searchers rarely avail themselves of advanced search features and boolean operators as a means to improve the precision of their queries and such facilities are used in less than 20% of searches and in more recent studies in less than 5–10% of searches.
(3) Searchers often look no further than the first page of search results.
(4) While early Web search activity focused on niche content areas (such as adult content), in time Web search has broadened to cover a wide variety of topics with entertainment and e-commerce services growing rapidly.

Of course all of the studies that we have included to date have focused on an analysis of what might be termed traditional Web search in the sense that they have all assumed conventional PC-based access. The world is changing and, as we have indicated previously, in the future many users are likely to spend a significant amount of their time online through alternative access channels such as their mobile phone. To date, there has been relatively little analysis of the behavior of mobile users as they use their mobile phones and PDAs to search for content, an issue that is central to the work described in this article. The question of whether the behavior of mobile searchers is different from their Web counterparts, given the usual device limitations, for example, is extremely important. Indeed understanding how mobile users are likely to search for information can help us to build better mobile search engines that are well adapted to the particular needs of the mobile Internet, its users, and its devices.

2.5 Google Mobile 2006

One recent study based on mobile search is presented in work by Kamvar and Baluja [2006] and covers an analysis of more than 1 million requests from Google's mobile search logs. In fact two separate logs were analyzed: Google's XHTML logs relate mainly to searches originating from mobile phone handsets with conventional input capabilities (e.g., a 12-key keypad); Google's PDA logs relate to searches originating on PDA-like devices with more sophisticated

input capabilities (and typically larger screen sizes). The study provides some insight not only into the world of mobile search, but also allows us to consider the impact of different device characteristics (mainly in relation to input/ouput capabilities). The study is a useful starting point and its results bear a strong resemblance (perhaps unsurprisingly) to those found for Web search, especially for some of the earlier studies.

For example, the average query length for XHTML queries (phone handsets) is 2.3, shorter than the 2.7 terms per query found for the PDA queries. Of course these queries require considerably more effort to input, especially on a mobile phone handset, than more conventional Web queries. This study also highlights the prevalence of adult-related searches. The Adult category was the top ranking category among the XHTML logs, suggesting that mobile search is still in the early stages of adoption where such interests tend to be favored. Whether we will witness the growth of alternative interest categories with the widespread adoption of mobile search as we have with the growth of Web search remains to be seen. The study also found a lower number of queries per search session (1.6 compared to > 2 for Web search studies), and a very low percentage of searchers were willing to look beyond the first page of results (8.5%) compared to Web search studies.[1]

2.6 Beyond the Google Mobile Study

On the surface, the Google study is a close fit with the sort of analysis that is needed to better understand mobile search behavior and shares a number of similarities with the work presented in this article. However, the Google mobile study suffers from a number of significant limitations, that are directly addressed by the study presented in this article. These limitations include the following.

(1) From the broad perspective of information access, the Google mobile study is limited by its exclusive focus on search as just one approach to mobile information access. The study remains silent on any comparison between search and alternative information access approaches such as browsing.
(2) The Google mobile study focuses exclusively on US searchers using Google's standard mobile search interface, which represents a reasonably direct translation of its Web search service and as such does not provide any special features that are designed to recognize or take advantage of the peculiarities of the mobile domain.
(3) A very significant limitation of the Google mobile study relates to its exclusive focus on XHTML content only and the fact that it ignores mobile-specific WML content. The significance of this limitation is highlighted by the fact that, in our own analysis, we have found that approximately 52% of requests were for WML content.

[1]It is worth mentioning here that the authors noted that this low figure may be a consequence of some interface confusion. They noted that approximately 32% of queries issued were duplicates leading them to conclude that users may be mistakenly re-searching instead of requesting the next page of results.

(4) The Google mobile study is based loosely on previous investigations into Web search but does not go far enough. For example, there is no detailed analysis of any extended search behaviors such as an analysis of query modifications. This type of analysis has proved to be important in the past when it comes to understanding how searchers locate content through a sequence of related queries by specializing and/or generalizing their original queries in order to refine their information needs.
(5) One of the most significant features of the mobile Internet is the vast array of devices in common usage. There are quite literally hundreds of different mobile devices on the market, many with very different mobile Internet features including a wide range of display characteristics and input capabilities. While the Google mobile study pays some attention to the relative effort associated with text entry on mobile devices in general terms, there is no explicit device-level analysis.

The work presented in this article extends and compliments the Google mobile study in a number of important ways.

(1) To begin with, the starting point for our study is a desire to understand mobile information access in broad terms, focusing on mobile search in particular, but also considering browsing as an alternative information access approach, and the difference between browsing and search behaviors.
(2) Our study more faithfully follows the methodology adopted by comprehensive Web search studies in the past (such as Silverstein et al. [1999], Jansen et al. [2000], and Spink et al. [2001]). To this end, we look beyond simple statistics, such as average query length, to include a detailed analysis of query modification strategies.
(3) Perhaps most importantly, the analysis contained within this article includes both WML and XHTML content and as such provides a far more comprehensive analysis of mobile information access, given that WML content accounts for approximately 52% of activity in our dataset.
(4) Where the Google mobile study focused on US users of a single mobile search engine (namely, the Google mobile search engine), our study analyzes the behavior of European users of more than 30 different mobile search engines, ranging from the larger players such as Google and Yahoo to a range of niche mobile search engines such as Click4Wap and TagTag.
(5) Explicit consideration is given, in our study, to an analysis of the various different types of devices used for mobile Internet access with particular attention paid to any differences in the devices used for mobile browsing versus mobile search.

3. RESEARCH QUESTIONS

In this study we are interested in answering a number of important questions, including some general questions about the information access habits of users and more specific questions about how individuals tend to search and what they are searching for:

(1) Is browsing still the dominant form of information access on the Mobile Internet?
(2) Are there characteristic differences between users who tend to search compared to users who rely solely on browsing?
(3) Is there an interaction between mobile devices and different modes of information access?
(4) How do users search for information on the mobile Internet?
(5) Are their query patterns similar to those reported in Web search?
(6) Do they avail themselves of advanced search features when available?
(7) Do they tend to modify their queries? If so, how?
(8) What types of information are mobile users looking for?

In the following, we will attempt to answer each of these questions in the context of the mobile Internet. In addition, where feasible, we will relate our answers to similar questions asked about the traditional Web and Web search in order to highlight any significant similarities or differences.

4. RESEARCH DESIGN

Our approach is based on an analysis of a significant mobile Internet transaction log from a major European mobile operator. This analysis allows us to investigate a number of important aspects of MI usage such as the relative importance of browsing versus search activity. In particular, it facilitates a detailed investigation of mobile search patterns including device and search engine usage, and a query-level analysis based on the methodology of Jansen et al. [2000] and Spink et al. [2001].

4.1 Data Collection

The data used in this study represents a collection of more than 30 million mobile Internet requests generated by just over 600,000 unique European mobile subscribers over a 24-hour period in late 2005. The requests were recorded in a transaction log. Each request within the transaction log includes the following information.

(1) *Anonymous User ID*: a unique (anonymous) identifier for each subscriber.
(2) *Timestamp*: The date and time (measured in hours, minutes, and milliseconds) for each request.
(3) *URL*: the URL of the page requested by the subscriber.
(4) *Status*: the HTTP status code corresponding to the result of the request, including, whether the request was successful, whether the user was redirected, etc.
(5) *Bytes*: the size of the requested page in bytes.
(6) *Device*: an identifier corresponding to the device used by the subscriber to make a given request (e.g. Nokia 6630, Siemens SK, Motorola V3, etc.)

Before starting our analysis, we needed to identify and extract (1) individual sessions and (2) individual search requests from the transaction log.

A *session* is defined as a sequence of actions performed by the same user. In our study, we identify sessions as a sequence of contiguous requests by a unique user delimitated by a 5-minute period of inactivity. Using the timestamp field and user id field, we identified just over 1.1 million individual user sessions within our dataset.

Each transaction within the log represents an individual request issued by an individual user. This request could be a *browsing request* where the user asks for a specific mobile page or it could a *search request* where the user submits a query to a search engine. The extraction of these search requests was quite challenging. Following are the steps we carried out to identify individual searches.

(1) We generated a set of *distinct hosts* by analyzing each URL associated with each tranaction/request within our log.
(2) From this list of distinct hosts, we identified and extracted a set of *distinct search services*.
(3) Next, we analyzed the query pattern per host to determine how search requests were formatted for each individual search service. The result of this step was a list of *search URLs*.
(4) Using regular expressions and pattern-matching, we were able to extract the *queries* from each of the search urls.

4.2 Data Analysis

To answer each of our research questions we explored the following key areas.

4.2.1 *Browsing vs Search.* We compared browsing versus search activities using two different sets of analyses. The first was a *session-level analysis* where we looked at various characteristics of mobile sessions including session duration, request length (measured in bytes), and session frequency. The second was a *user-level analysis* where we compared and contrasted mobile users who exclusively browse for information to those users who have adopted search as a complimentary route to interesting content. Section 5.1 and Section 5.2 describe the results of these two analyses.

4.2.2 *Search Engines and Devices.* After extracting the complete set of search URLs from our data (see the previous section for further details), we were able to examine the relative popularity of these mobile search engines. We identified over 30 mobile search engines within our dataset including major players like Google and Yahoo as well as independent WAP search engines like Click4Wap and Mooobl. We also identified some specialized search services like e-commerce search services and ringtone search. The relative popularity of these search engines was determined by examining the percentage of searches issued to each engine. See Section 5.3.1 for full details.

We also consider the different mobile devices common among mobile searchers compared to mobile browsers. To carry out the device analysis, we extracted the top-10 browsing devices and the top-10 search devices from our dataset. We then conducted some research into each of these devices, focusing

on key features like screen size, input-options, and Internet/browsing support. We were then able to compare each of the devices using these feature sets. See Section 5.3.2 for full details.

4.2.3 *Emerging Search Behavior.* We performed detailed query and topic analyses to investigate how people are searching for information and what they are searching for on the mobile Internet. A *query* within our dataset is defined as a string of terms submitted by a mobile user to one of the 30 mobile search engines identified. A *term* is defined as a series of characters delimited by a white space. After extracting the complete set of queries from each of the search URLs within our dataset (see Section 4.1 for details of how we extracted these queries), we were able to report basic statistics like query length, query frequency, and use of advanced features as well as more detailed query-level analyses like query modifications, query repetition, and popular query topics. See Sections 5.4 and 5.5 for full details.

5. RESULTS

In this section, we report the results of our large-scale study of mobile information access patterns.

5.1 Session-Level Analysis

As discussed, we can usefully consider two basic modes of information access, browsing and search. While search now dominates the Web, browsing has traditionally dominated the mobile Internet. The reason is that the mobile Internet is better adapted to browsing in a number of important ways. For a start, most subscribers access the mobile Internet through their operator's portal, for example, portals such as Vodafone Live! and O_2 Active are leading European access points. Second, mobile devices offer users a number of features that are clearly designed with browsing in mind, for example, most MI-ready handsets will include special-purpose browsing keys to help users navigate through portal menus. Third, data entry is not straightforward on most mobile handsets making it difficult for users to enter search queries. Finally, until relatively recently, most operators offered very limited search facilities (if at all) from their portals.

In this, the first part of our study, we will analyze the relative importance of browsing-based versus search-based mobile information access. Specifically, we wish to understand whether the recent introduction of search-based services is leading to an increase in search-based activities. Recent evidence suggests that search activity is gaining some momentum on the mobile Internet, but exactly how common is this type of information access behavior and how important is it likely to become? As we shall discuss, sessions that involve mobile search engine usage are now making a significant and growing contribution to overall mobile Internet traffic. With significant growth in the quantity of mobile-specific content, search facilities are becoming more commonplace. In particular, it is worth highlighting how search plays a critical content discovery role in the mobile Internet by allowing subscribers to access off-portal content.

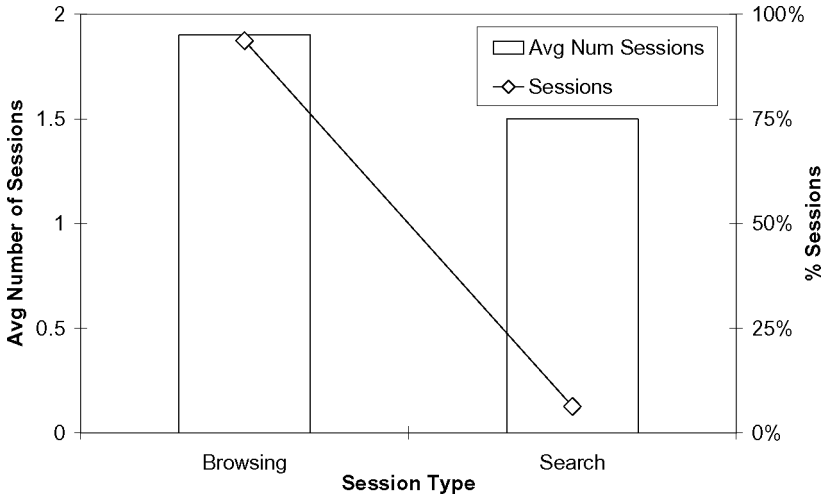

Fig. 1. Average number of sessions and percentage of sessions per-user per-day.

By comparison, browsing tends to limit the user to their home (operator-owned) portal.

To begin with, we define two types of session.

(1) *Browsing Session.* A browsing session refers to a session in which there is no identifiable search activity among the requests. In other words, the user has accessed content by following a series of links probably through the operator's portal or from some previously bookmarked page.
(2) *Search Session.* In contrast, a search session is a session where the user has engaged in at least some search activity. By this, we mean that they have entered at least one query to a known mobile search engine.

Importantly, while browsing sessions by definition contain no search-related activities, the converse is not true for search sessions, that is, search sessions can, and usually do, contain browsing-based activities as users search and then browse for content. Users will usually search for an interesting site and then browse for content within the site, for example.

When we examine the 1.1 million sessions in the test data, we find that 94% are browsing sessions, that is, over 1 million sessions contained no search-related activity. Clearly, browsing continues to dominate in the portal-based world of the mobile Internet. However, the recent availability of search services is beginning to have a recognizable impact with more than 73,000 (6%) search sessions evident within the test data. In other words, 73,000 mobile Internet sessions included at least some search-related activity. Moreover, as we will discuss, although the relative percentage of search sessions remains modest, these sessions make a significant overall contribution to MI usage in terms of their average request length, bytes downloaded, and session duration.

Figure 1 compares the number of browsing and search sessions per-user per-day among those users who engage in such activities. We see that the average number of search sessions per-user per-day is 1.5, while the average number

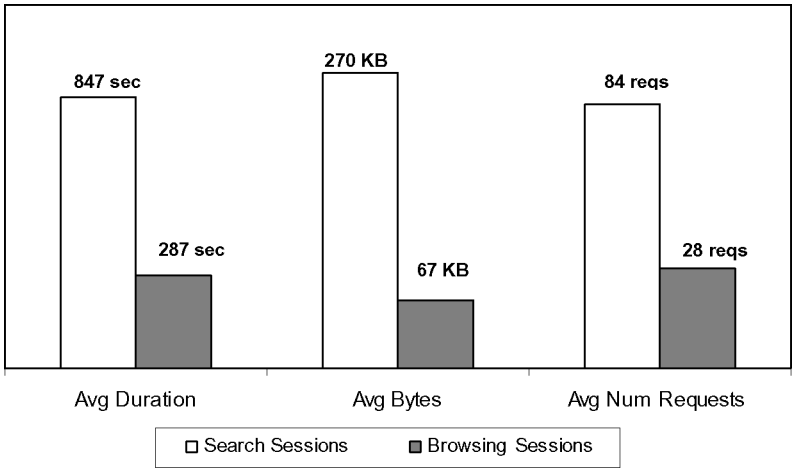

Fig. 2. A comparison between daily search and browsing sessions in terms of session duration, bytes downloaded, and numbers of requests.

of browsing sessions per user is 1.9. Therefore, users engage in more browsing sessions than search sessions. However, this relative frequency of search usage is only a very crude measure of search-related activity and, ultimately, it is important to understand the extent to which mobile search has the potential to engage mobile users in more meaningful mobile sessions. For instance, mobile search has the potential to help users to discover entire new domains of off-portal mobile content and as such may reasonably be expected to serve as a catalyst for extended mobile Internet sessions if such content is found to be of interest to the user. With this in mind, we also examined how browsing and search sessions differed in terms of their typical length (the average number of requests), size (the average bytes downloaded), and duration (seconds).

The results presented in Figure 2 are striking and indicate a very significant difference between both session types. For example, search sessions are found to be longer (in terms of requests and duration) and larger (in terms of bytes downloaded) than browsing sessions. For instance, search sessions are nearly 3 times the duration of browsing sessions, comprise more than 2.5 times the number of requests, and lead to 4 times the number of bytes downloaded.

These results suggest, at the very least, that search is helping users to locate content that is sufficiently interesting to entice them to stay online longer and download more content. Thus, as mentioned earlier, although search sessions represent a small proportion of total sessions, they make a disproportionately large contribution in terms of their relative length, size and duration when compared to browsing sessions. For example, as shown in Figure 3, the 6% of search sessions are responsible for 17% of the total requests, 23% of the bytes downloaded, and 17% of the total session time in our test data. Indeed given that many mobile operator charging-models are based on bytes downloaded, it becomes quite clear how valuable search-related activity can be since each additional search session can deliver 4 times the bytes of a typical browsing session, and thus up to 4 times the revenue for the mobile operator.

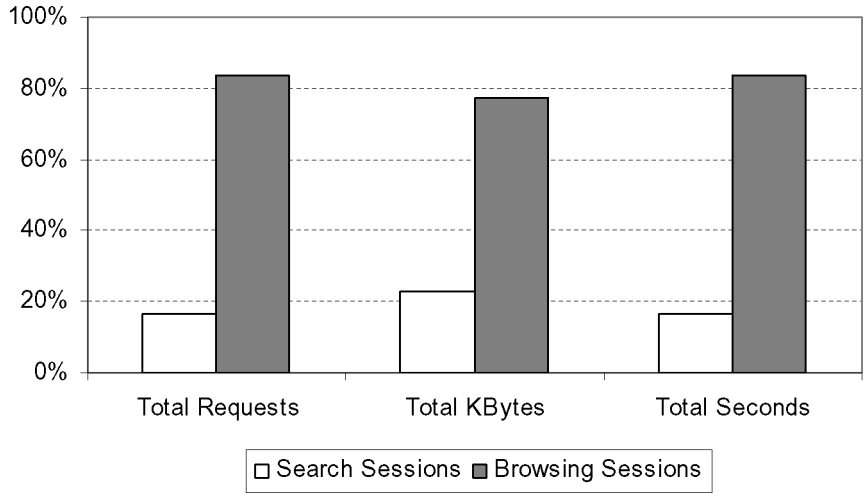

Fig. 3. A comparison of the relative contribution of search and browsing activity in terms of requests, bytes, and time online.

In relation to these differences between search and browsing sessions, it is, of course, important to remember that search sessions are not exclusively concerned with only search activities. As defined previously, a search session is any session that includes at least some search activity, but also typically includes significant browsing activity. After all, browsing is a common result of successful searching because the user locates an interesting site or service which they then browse through to locate particular relevant content. It is also worth highlighting that searching is an inherently more challenging information access task. When users browse for content, they are often navigating to information that they have found in the past and so are more confident when it comes to relocating the information again. In contrast, many searches are devoted to locating new information and will typically involve more effort on the part of the searcher. Certainly as we will discuss in future sections, it is clear that there are significant problems facing mobile users when it comes to searching effectively. For example, we will discuss how mobile searchers often need to modify their initial search queries in order to find relevant results, more so than in traditional Web search. This suggests that mobile searchers have to work harder to locate relevant information than their Web counterparts and is also partially responsible for the longer search sessions compared to browsing sessions.

5.2 User-Level Analysis

A session-based analysis allows us to draw some general conclusions about the similarities and differences between different types of sessions, independent of the users who engage in these sessions. But understanding the users is equally important. Specifically, we might ask whether there are differences between users who search compared to those who browse?

We have previously suggested that mobile search plays a key off-portal content discovery role, helping users to access a growing universe of mobile

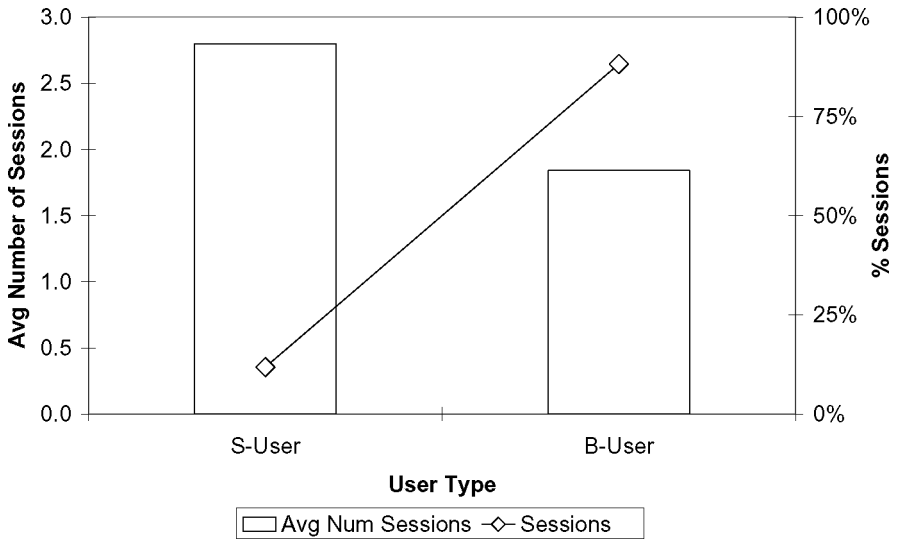

Fig. 4. Average number of sessions and percentage sessions per-S-User and B-User per-day.

content rather than limiting their access to browsing the 'walled-garden' of the operator's portal. If this is the case, then we might legitimately expect to find significant usage differences between users who search and users who browse. For example, we might expect to find an increased level of usage among users who search (if their searches tend to lead to interesting and relevant content beyond the portal). Or perhaps instead, we will find that there is no such difference, that browsing sessions are usually just traded for search sessions without contributing to an increase in usage at the user level.

In an effort to answer these questions, we first define a *search user* (S-User) to be any user who has engaged in at least one search session; thus a search user is a user who, at some level at least, recognizes the value of mobile search. Of course the sessions that a search user has engaged in will tend to include search sessions and browsing sessions; sometimes they will browse for information, other times they will search. Conversely, we can define a *browsing user* (B-User) to be any user who has not engaged in any search sessions; their sessions will, by definition, only include browsing sessions.

We can now compare search users to their browsing counterparts. To begin with, while we previously found only 6% of total sessions to be search-related, we find that 8% of our users (just under 50,000 users) are search users. More revealing again are the average daily usage figures for search and browsing users in terms of their average sessions, requests, bytes, and time online per day. In Figure 4, we see that the average number of sessions per-search-user is 2.8, while the average number of sessions per-browsing-user is only 1.8, that is, search users engage in 55% more daily sessions than their browsing counterparts. On average, for these 2.8 daily sessions, approximately half are search sessions, while the other half are browsing sessions.

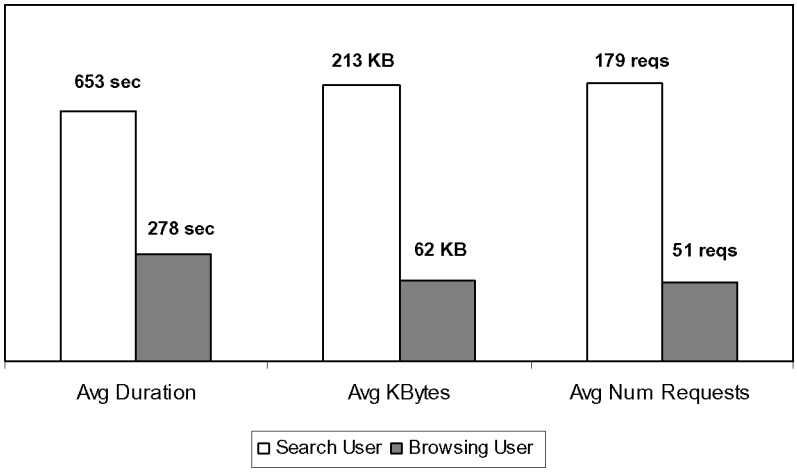

Fig. 5. A comparison of S-User and B-User sessions in terms of requests, bytes, and seconds online.

Even more significant differences between these two user types become apparent when we look at the average number of requests, time online, and bytes downloaded per day (see Figure 5). For example, the average search user generates more than three times as many requests per day than the average browsing user (179 vs. 51 requests each, respectively), downloads more than three times as much content (213kb vs 62kb, respectively) and spends more than twice as long online than the average browsing user (653 seconds per day vs 278 seconds per day, respectively).

Clearly searchers, although small in number (at least relative to the more traditional browsers), are responsible for a disproportionately large volume of online usage: 23% of the total requests generated across our usage logs are generated from the 8% of users who engage in search-related activities. That said, the direction of the arrow of causality is far from clear in this instance. Are search users naturally more active and thus more likely to engage in search-related activities? Or is it the very act of using mobile search engines that inevitably leads to this increased level of activity? Probably it is a combination of both. We do know from the previous section, for example, that sessions which contain search activity are much richer than sessions that do not. But at the same time, it is possible that today's searchers correspond to a particular type of user with specific information needs that can only be addressed by search. We will return to this issue in future sections and argue that there is strong evidence to support this particular hypothesis especially when we look at the type of content that users tend to search for.

5.3 Mobile Search Engines and Device Usage

With mobile search on the increase, it is worth considering the search engines that subscribers are relying on and the devices that they are using to perform their searches. In this regard, the mobile Internet differs quite significantly from the traditional Web. With mobile search still in its infancy, there

Table I. Mobile Search Engine Usage in Order of Popularity

Rank	Search Engine	% of Searches
1	Google	76.2%
2	Operator-Specific Search	14.4%
3	Yahoo	5.7%
4	Independent Wap Search Engines/Directories *(Includes: TagTag, Click4Wap, Mooobl, Ithaki, Seek4Wap, Freowap, WapAll, WapMeter, etc)*	1.6%
5	eCommerce Search	1.6%
6	Other Major Search Engines *(Includes: AskJeeves, AllTheWeb, MSN, AOL, AltaVista, Excite, Nokia and HotBot)*	0.4%
7	Misc *(Includes: image-specific search, ringtone-specific search and other misc search services)*	0.1%

is a fragmented collection of potential search engines and services available to users. And given the wide range of mobile devices on the market, it is also interesting to consider whether the same or different devices are found to be used during browsing and search sessions.

5.3.1 *Search Engines of Choice.* At the time of writing, the Web search market is dominated by Google (42% of searches), Yahoo (28%), and MSN search (13%), with AOL and Ask capturing just 8% and 6% of searches, respectively [ZDNet 2006a]. Thus less than 3% of searches are handled by search engines other than those listed. In this section, we consider whether a similar pattern of usage exists within the mobile search space.

In total, the 73,000 search sessions available in our study represented over 400,000 individual searches (see Section 5.4 for a more detailed query-based analysis). These searches used more than 30 different mobile search engines, ranging from major services such as Google and Yahoo to independent WAP search engines such as TagTag, Click4Wap, and Mooobl, as well as operator-specific search services (offered through their portals) and specialized search services such as e-commerce services and ringtone search. As Table I shows, while Google remains the leader in mobile search (attracting more than 76% of the queries), independent mobile search engines make up approximately 20% of mobile search. Other than Google and Yahoo, the traditional Web search giants have been slower to address the need for search on the mobile Internet. It is interesting to note that operator-specific search accounts for just over 14% of mobile search. This suggests that our test users were often ignoring the search services made available as part of their operator's portal in favor of Google.

5.3.2 *Device Usage.* The mobile Internet attracts a wide range of devices offering a variety of different features including different display characteristics and input options. For example, it is not unusual for a mobile operator to have to consider the features of upwards of 50 different mobile device groups, each group requiring a separate set of stylesheets so that content may be optimized for the set of devices in that group. This level of device optimization far exceeds that which is the norm on the Web where perhaps only a few

basic browser types/versions need to be considered[2]. In this section, we consider whether there are specific devices that appear to be better adapted to search or browsing-related activities. To do this, we focus on the top-10 most popular devices used in browsing sessions and the top-10 used in search sessions, and we characterize these devices in terms of the following features. (Note that we have labeled the devices rather than providing actual device names for anonymity.)

(1) **Interface/Display**
Display Type: The type of screen, for example, TFT, CSTN, etc.
Display Size: The size of the display in pixels.
Colors: The number of colors supported by the display.

(2) **Internet/Browsing**
WAP: The version of WAP supported.
XHTML: Is XHTML browsing supported?
Email: Is email supported?

(3) **Input Capabilities**
Predictive Text: Does the device support predictive text input? Predictive text provides a type of text completion for more efficient text input.
Navigation Control: The availability of special purpose navigation keys to assist with navigation and menu selection.

(4) **Multimedia**
Camera: Indicates the presence of a built-in camera.
Music: Indicates whether the phone can store and play music files (e.g., AAC, MP3, etc.)
Radio: Indicates whether the phone includes integrated FM radio.
Video: Indicates whether the phone can capture and playback video clips.

Table II presents the top-10 mobile devices used in browsing sessions. These handsets account for 42% of the browsing requests generated in browsing sessions. Perhaps as expected, many of these devices are modern handsets. Most support color interfaces, predictive text entry, and XHTML browsing. All ten also include some type of navigational control to support online browsing and navigation. However, it is worth noting that many of the devices have very limited display characteristics. In fact, 3 out of the top-5 devices have very small (128×128) displays and two of these offer only limited color resolution; these two devices are also restricted to WAP 1.2.1 and, therefore, cannot reap the benefits of WAP 2.0 browsing [Open Mobile Alliance 2002].

Table III presents the top-10 mobile devices used in search sessions; these devices cover 56% of all searches generated. It is interesting to note that there are differences between the devices that are popular for browsing and those popular for search. In particular, there is a clear tendency towards more sophisticated devices when it comes to search. The majority of the popular devices used in search benefit from high-resolution color displays; in fact, 5 of the

[2]Incidentally, this is why device management middleware is so important in the mobile sector.

Table II. Top-10 Mobile Devices Used in Browsing Sessions

Number	Mobile Device	% of Usage	Display Type	Display Size	Colors	WAP (Version)	XHTML	Email	Predictive Text	Navigation Control	Camera	Music Player	Radio	Video
1	A	9%	TFT	128 × 128	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
2	B	7%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	-	x
3	C	6%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	x	x
4	D	4.1%	CSTN	128 × 128	4K (12-bit)	1.2.1	x	x	x	x	x	-	x	-
5	E	3.6%	CSTN	128 × 128	4K (12-bit)	1.2.1	x	x	x	x	-	-	-	-
6	F	3.1%	TFT	176 × 220	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
7	G	2.9%	TFT	208 × 208	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
8	H	2.7%	TFT	170 × 220	262K (18-bit)	2.0	x	x	x	x	x	-	-	x
9	I	2.3%	TFT	128 × 128	65K (16-bit)	2.0	x	x	x	x	x	-	-	x
10	J	2.2%	TFT	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	-	-	-

Table III. Top-10 Mobile Devices Used in Search Sessions

Number	Mobile Device	% of Usage	Display Type	Display Size	Colors	WAP (Version)	XHTML	Email	Predictive Text	Navigation Control	Camera	Music Player	Radio	Video
1	C	17.6%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	x	x
2	F	7.4%	TFT	176 × 220	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
3	B	6.5%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	-	x
4	A	5.9%	TFT	128 × 128	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
5	K	4.8%	TFT	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	-	-	-
6	L	3.9%	CSTN	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	-	-	-
7	H	2.6%	TFT	170 × 220	262K (18-bit)	2.0	x	x	x	x	x	-	-	x
8	M	2.5%	TFT	240 × 320	262K (18-bit)	2.0	x	x	x	x	x	x	x	x
9	N	2.5%	TFT	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	x	-	x
10	O	2.4%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	x	x

top 10 support 18-bit color and (relatively) large screen displays (176 × 220). They all support WAP 2.0, and they all provide navigation controls.

Overall we found that the devices used across both browsing and search sessions were standard mobile phones, all quite modern and most with significant screen space, color interfaces, predictive input, and XHTML support. In fact, we found that 5 of the devices listed, namely, A, B, C, F, and H, were present in the top-10 lists for both search and browsing sessions. Despite this, our device analysis has shown that the phones used to search are generally more sophisticated that those used to browse. For example, 50% of the phones in Table III have highly sophisticated displays particularly in relation to screen size and resolution. Interestingly, we found no significant differences in relation to input capabilities between the two device sets. None of the phones listed in either

table have flip-out keyboards or stylus support, while all of the phones listed have predictive text. Although it is unlikely that mobile information access is at the top of someone's preferences when purchasing a new mobile phone, it is reasonable to assume that users with higher spec phones are more likely to use the mobile Internet or at least find it easier to use if they venture online via their mobile phone.

5.4 Query-Level Analysis

So far we have compared search and browsing behavior on the mobile Web from a relatively high level of abstraction, focusing on sessions and users, search engines and devices. In our current research, we are especially interested in understanding more about the nature of mobile search—how mobile subscribers actually search, the terms they use in their queries, and the way that they form these queries—so that we can, in the future, develop search solutions that are better adapted to the needs of mobile subscribers. Thus in this section, we look at the search queries submitted during search sessions. We base our analysis in large part on the methodology adopted in Jansen et al. [2000] and Spink et al. [2001], which analyzed the search behavior of Excite users during March 1997 and September 1997, respectively; we will refer to these as the Excite 1997a and the Excite 1997b studies. Conveniently, both of these studies also rely on a single day of log data and so serve as a useful basis for comparison when it comes to evaluating the results of our mobile search analysis. Will there be a strong similarity between the observations made by these analyses of early Web search and our observations for emerging mobile search? We might expect strong similarities between the behavior of Web users in these early Web studies and today's early adopters of mobile search. Or alternatively, we might find significant differences between our mobile searchers and their Web search counterparts indicating some fundamental differences between how users search on the Web and the mobile Internet.

5.4.1 *Summary Data.* As mentioned already, our search data consists of more than 400,000 separate search queries spanning just over 73,000 search sessions generated by almost 50,000 users. Table IV presents a variety of statistics based on an analysis of the queries and terms used as part of these searches, presented side-by-side with comparable statistics that we have compiled from the two early Excite studies[3]. There are a number of features worth noting.

(1) Mobile search queries are short and, with 2.06 terms per query, they are somewhat shorter than their Web counterparts. Indeed, there is evidence that today's Web search queries are now reaching an average of 3 terms per query [ZDNet 2006b].
(2) Mobile searchers rarely use advanced search features, with only 3% of mobile queries containing any of the standard query modification operators

[3]Some statistics associated with the Excite 1997a dataset were not explicitly listed in the corresponding research paper [Jansen et al. 2000]. Therefore, we had to extract this information from the relevant text in the paper. A value labeled of unavailable in Table IV illustrates a result that we were unable to extract.

Table IV. A Comparison of Summary Statistics for Mobile Search and Early Analyses of Web Search

Variables	*Mobile (2005)*	*Excite (1997a)*	*Excite (1997b)*
Number of Users	48933	18113	211063
Number of Sessions	73289	Unavailable	211063
Mean Sessions per User	1.5	Unavailable	1
Number of Queries	423764	51473	1025910
Number of Unique Queries	91989	29347	531416
% Unique Queries	22%	57%	52%
Number of Repeat Queries	324115	22127	395461
% Repeat Queries	76%	43%	39%
Number of Zero-Term Queries	7660	2584	99033
% Zero-Term Queries	2%	5%	10%
Number of Terms	872926	113793	2216986
Number of Unique Terms	40341	21862	140279
% Unique Terms	5%	19%	6%
Mean Queries per User	8.66	2.84	4.86
Mean Queries per Session	5.78	Unavailable	4.86
Mean Terms per Query	2.06	2.21	2.16

(and, or, quotation marks, etc.). Web searchers more frequently take advantage of advanced search features with up to 20% of searches containing boolean operators or query modifiers in the Excite 1997b dataset [Spink et al. 2001].

(3) Our early-adopter mobile searchers appear to engage in significantly more searches per session (8.66) than the Excite 1997b Web searchers (4.86 searches per-user per-session) [Spink et al. 2001] and the Web searches from the Excite 1997a study (2.84 searches per-user per-session) [Jansen et al. 2000], suggesting perhaps that it is more difficult to locate relevant information in mobile search.

(4) There is a relatively low percentage of unique queries among the mobile searches (22%) compared to Web search (57% and 52% from the 1997a and the 1997b data, respectively), suggesting a more limited mobile search vocabulary.

(5) Conversely, we find a much higher incidence of repeat queries in mobile search (76%) compared to Web search (43% and 39% for 1997a and 1997b data, respectively).

Of course these summary data only tell part of the story and the averages in particular are based on skewed underlying data. The following sections take a closer look at the distribution of these data to shed more light on mobile search behavior.

5.4.2 *Query Length.* As already stated, mobile searchers tends to use queries that are marginally shorter (mean terms per query = 2.06; standard deviation = 1.13) than their Web counterparts. For instance, in addition to average Web search query lengths of 2.16 and 2.21, as reported in Table IV, Silverstein

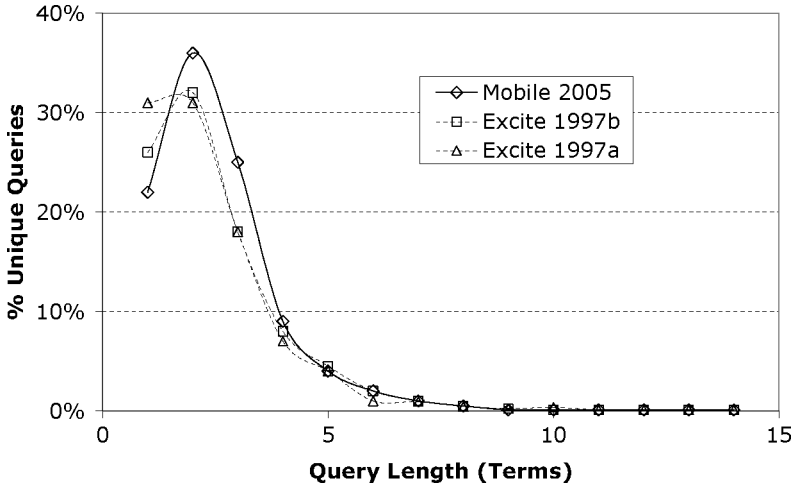

Fig. 6. Percentage of unique queries of different lengths.

et al. [1999] report an average length of 2.35 (standard deviation = 1.74) and recent statistics from Yahoo suggest that average number of terms per query in Yahoo's search offering has grown to 3 [ZDNet 2006b]. Interestingly, recent mobile search statistics produced from an analysis of Google's mobile search service [Kamvar and Baluja 2006] highlights an average of 2.3 terms per query, slightly longer queries than those found across the 30 search engines in our study, suggesting a tendency for Google's mobile searchers to use slightly longer queries than the non-Google mobile searchers. This hypothesis is further supported by our own mobile data. When we focus on search queries submitted just to Google in our dataset, we find an increased query length of 2.22 terms (standard deviation = 1.15) compared to 1.56 terms per query for non-Google searches.

Examining the query lengths of our full set of mobile search queries in more detail reveals a distribution of term lengths that is almost identical to that found for Web search. This suggests that mobile searchers are searching for content in broadly similar ways when compared to early Web search; their shorter queries are likely to be a result of the more restricted text-entry features of mobile devices. For example, Figure 6 presents the percentage of mobile queries with different term lengths (in the range 1 to 15 terms), alongside the results found for early Web queries by Jansen et al. [2000] and Spink et al. [2001]. Approximately, 58% of mobile search queries contain two terms or less, a figure that is almost identical to the distribution of query lengths in the Excite 1997b Web data and slightly lower than the 62% of queries with these lengths in the Excite 1997a Web search data. Ultimately, about 85% of mobile queries have 3 terms or less; this figure is slightly larger than that found for early Web search results where 76% (Excite 1997b data) and 80% (Excite 1997a) of queries had 3 terms or less.

5.4.3 *On the Use of Advanced Search Features.* So-called advanced search features (e.g., the availability of logical operators and other query modifiers)

Table V. On the Use of Advanced Search Features Among Mobile Queries (Total Number of Queries = 423,764)

Feature	Number of Queries	% of Queries
AND/And/and	2,912	0.68
OR/Or/or	84	0.02
NOT/AND NOT	3	0.0007
+ (plus)	1,654	0.39
− (minus)	6,843	1.6
“” (quotations)	1,085	0.26
() (parenthesis)	122	0.03
&	613	0.14
Total	12,910	3.04 %

have long been a feature of Web search, largely as a result of the IR origins of Web search engines. Interestingly, while expert users of traditional IR systems put such operators to good use, the same is typically not true in Web search. For example, Spink et al. [2001] found that less than 10% of early Web search queries made any use of such advanced search features: boolean operators were used by less than 5% of the queries (with the logical AND being the most common boolean operator); term modifiers (e.g., ‘+’ and ‘−’) were used in about 7% of queries; and quotation marks were used to create phrases in 5% of the queries. The Altavista 1998 study [Silverstein et al. 1999] found an even greater use of query operators among Altavista users during 1998, reporting how more than 20% of search queries contained at least one query operator.

Our analysis suggests that the use of such features is even more rare in mobile search. As Table V indicates, in total, only 3% of queries included any special advanced search features at all with boolean operators being used in less that 1% of all queries. The cause is likely to be associated with the text-input difficulties associated with mobile devices and the use of such features usually demands special characters that are often difficult to access via a typical handset keyboard.

To relate back to our device analysis in Section 5.3.2, we found no correlation between the keyboard size/input capabilities of the devices used to execute these advanced queries. That is, the majority of the devices used to execute these advanced queries were standard mobile phones with no special input capabilities. Most modern mobile phones enable users to access a set of special characteristics with a single key-press followed by some scrolling to access the character in question, and it is likely that most of the advanced queries were executed in this manner.

5.4.4 *Queries Per User.* Figure 7 shows the number of unique queries submitted by users and demonstrates a strong correlation between mobile search and early Web search. The distribution of queries is highly skewed toward the lower end of the number of queries submitted per-user per-day with almost 50% of mobile searchers submitting just one query per day; this is similar to the Excite 1997b Web search data [Spink et al. 2001] but lower than that found for the Excite 1997a Web search data [Jansen et al. 2000]. To look at this another way, the mobile search data and the Excite 1997b Web search data show that

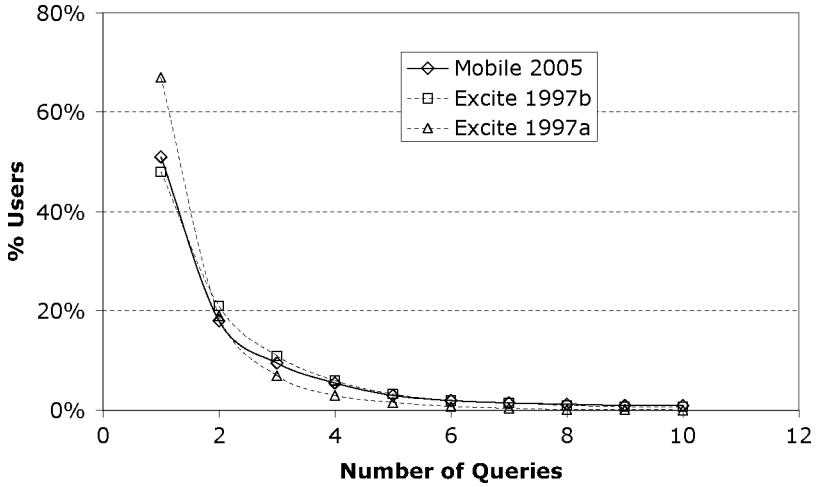

Fig. 7. The percentage of users submitting different numbers of unique queries per session.

20% of users engage in at least 4 searches per day, compared to only 7% of users in the Excite 1997a Web search logs.

5.4.5 *Query Modifications.* It is also interesting to look at the number of queries used per search session on the basis that sequences of queries within an individual session are likely to reflect the searcher attempting to locate some piece of information by modifying the original query in different ways. The distribution of mobile queries per session in our mobile data is presented in Figure 8 where we see that only 45% of search sessions contain a single query, and approximately 18% of sessions use 4 or more queries. One interpretation of this is that, at best, only 45% of searches lead to the target information in a single search, that is, with the submission of a single query, and that, conversely, more than half of the time searchers must use two or more queries before they find what they are looking for. Of course, these percentages represent upper bounds on search success rates only. There is no guarantee that just because a user exits their session after a single query that they have found their target information in the result list returned.

The Excite studies do not measure this statistic directly, they focused on queries per user only rather than queries per session and so cannot be reliably used as a direct source of comparison for our mobile search data. However, the Altavista 1998 [Silverstein et al. 1999] study did publish queries-per-session data for Altavista's Web searchers. This data is shown for comparison purposes in Figure 8 where we see a distribution for early Web search that is more skewed towards lower numbers of queries per session than our mobile search data. This suggests a potentially higher search success rate for Altavista's 1998 searches; almost 78% of search sessions involved a single query, for example. Indeed, only 9% of early Web search sessions required the use of more than 2 queries compared to 26% of mobile search sessions.

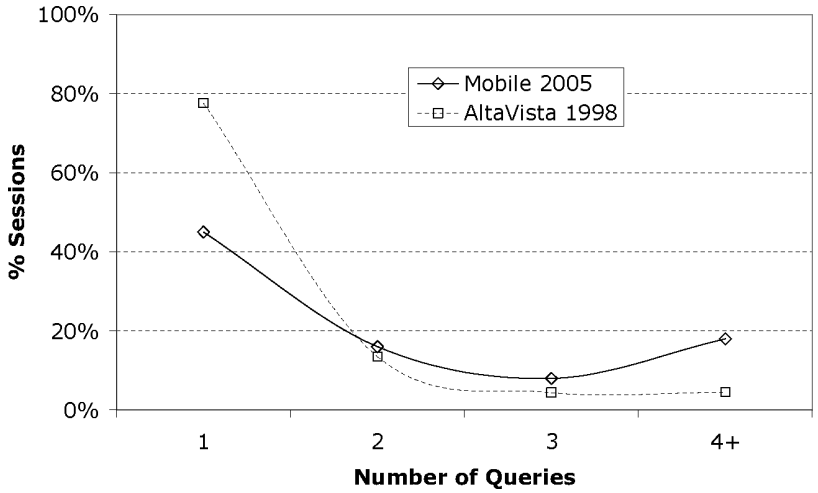

Fig. 8. The percentage of mobile search sessions with different numbers of queries.

These statistics highlight how sequences of search queries are commonplace, especially in mobile search, and they suggest that many users incrementally modify their original query in order to adapt or refine their search. To investigate query modification in more detail, we can classify search queries into one of a number of different types, according to the methodology adopted by Spink et al. [2001] and Jansen et al. [2000] (see Table VI)

—*Initial queries* correspond to the first query that occurs in a search session. The Excite 1997a study [Jansen et al. 2000] refer to these as unique queries which we feel is somewhat ambiguous and hence our change in terminology.

—*Modified queries* are any subsequent queries entered by the same user as part of a given search session. By defining modified queries in this way, we are assuming that all queries in a given search session correspond to the same general search objective. Nevertheless, it is a reasonable assumption that has been shown to hold in many search sessions, and it has been used in prior studies of this nature [Jansen et al. 2000].

—*Identical queries* are queries that occur during a search session but that are the same as the corresponding initial query. Because of the way in which we analyze our search data, the vast majority of in-session identical queries correspond to a user selecting the next page of search results for a given query rather than signaling the reentry of an identical query.

—*Zero-term queries* correspond to searches that are initiated with empty queries.

Perhaps the first thing to notice is that only 40% of queries correspond to the user initiating or refining a search. This compares with 57% in the 1997a Excite data [Jansen et al. 2000]. For example, in our mobile data, we find that only 17% of the queries are initial queries signalling the start of a new search, compared to 35% and 48% in the Excite datasets [Jansen et al. 2000; Spink et al. 2001]. So while similar degrees of query modification appear to occur (23% of

Table VI. Initial, Modified, Identical and Zero-Term Queries

Query Type	Number of Queries	% of Queries
Initial	72,656	17%
Modified	97,893	23%
Identical	245,555	58%
Zero-term	7,660	2%

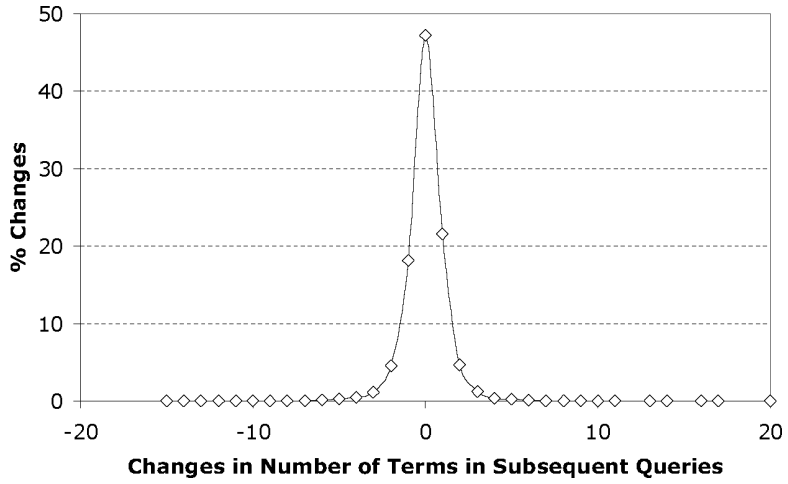

Fig. 9. Patterns of query modifications.

queries are modified in the mobile data and 22% are modified in the Excite 1997a data), mobile search is characterized by a lower proportion of initial queries and a higher proportion of identical queries. This is likely a side-effect of the limited screen space for result lists on the mobile device. In Web search, the norm is for 10 results to be displayed per page. In mobile search, it is usual for fewer results to be displayed per page, and thus a mobile searcher will have to request additional result-pages to receive a similar number of results as a Web searcher, hence the increase in identical queries which correspond to such additional result page requests in our logs.

To investigate how searchers are modifying their queries within a session, we carried out a similar analysis to Jansen et al. [2000], that is, we counted the change in the number of terms between subsequent queries. For this analysis, we concentrated on the 97,893 modified queries. A change of zero indicates that the user modified one or more terms but that the total query length remained the same, that is, the user swapped one term for another term without changing the query length. An increase of one term means that one term was added to the preceding query, and a decrease of one term means that one term was subtracted from the preceding query.

We can see from Figure 9 that in 47% of modified queries, there was a modification of one or more terms, but there was no change in the query length. This means that in almost half of the modifications, the modified query involved the user simply swapping one term for another term, and the resulting modified

query had the same number of terms as the preceding query. After examining this specific type of modified query, we noticed that some of these modifications were because of misspellings but most of these types of modifications represented the user attempting to direct or focus the query slightly without adding more terms. When we examined the remaining modified queries where terms were either added or subtracted, we saw that 28% of queries involved the addition of terms with the user specializing their query in some way, while 25% involved the subtraction of terms where the user generalized their query in some way. In summary, 22% of modified queries have one more term than the preceding query and approximately 18% have one less term.

Comparing these statistics to the modification behavior of early Web searchers [Jansen et al. 2000], we find some interesting differences. First, only 32.5% of Web query modifications were length invariant. So mobile searchers seem much more likely to modify queries by replacing or swapping terms rather than by adding or removing terms when compared to the early Web search counterparts. Moreover, we also notice a much more even distribution of term addition and deletions (22% and 18%) in mobile search compared to early Web search. For example, Jansen et al. [2000] reported that more than 41% of query modifications resulted in terms being added to the query (more specialized queries) and just under 26% of modifications subtracted terms from the query (to produce more generalized queries), suggesting that early Web searchers had a much greater tendency to move from broad to narrow queries.

There has been some interesting work carried out in the query recommendation area which could be exploited in this regard. This degree of query modification suggests that mobile users could benefit from having alternative queries recommended to them instead of generating the query modifications themselves.

5.4.6 *Query Overlap and Similarity.* In this section, we examine query overlap or query similarity across the queries. Using a simple measure of query overlap (also known as query similarity) taken from Balfe and Smyth [2005], we can measure various degrees of overlap between queries, ranging from duplication of terms to partial overlaps. To measure the overlap between two queries, we compute the degree of overlap between their query terms using Equation (1). For example, if we take the queries "free ringtones" and "new ringtones," we can see that they have one term in common out of three unique terms. Therefore, the overlap between the two queries is 0.33.

$$Sim(q_T, q_i) = \frac{|q_T \cap q_i|}{|q_T \cup q_i|}. \tag{1}$$

Figure 10 shows the percentage of queries that have overlaps/similarities above a certain similarity threshold. We have graphed these percentages for each query using five different similarity thresholds: > 0, ≥ 0.25, ≥ 0.5, ≥ 0.75, and 1. A similarity threshold of 1 indicates a duplicate query but not necessarily a query with all the terms in order.

The graph illustrates a high percentage of overlap at the $>$ 0 and $\geq$ 0.25 thresholds. Approximately 82% of queries share at least one term and

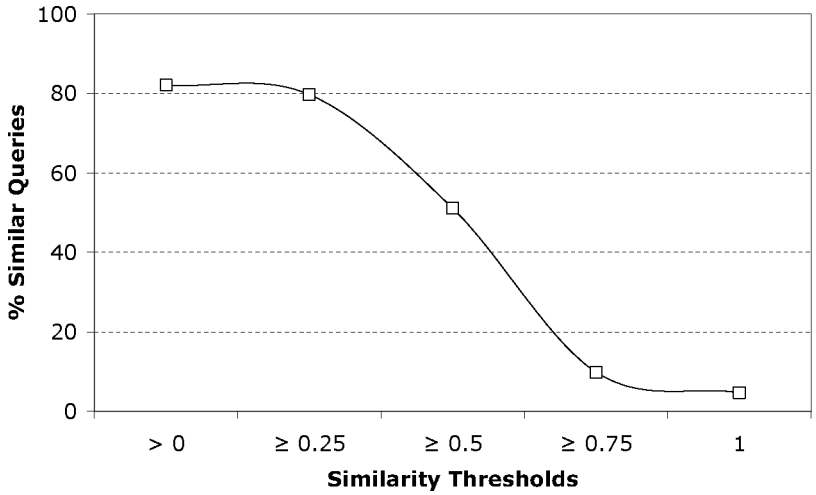

Fig. 10. Query overlap: the percentage of similar queries at different similarity thresholds.

approximately 80% of queries share at least 25% of their terms. When we examine queries that share half their terms, that is, ≥ 0.5 threshold, we see that over 50% of queries fall into this bracket. However, the degree of overlap falls sharply at the 0.75 threshold. Approximately 10% of queries share 75% of terms, and this decreases to 5% for duplicate terms (i.e., a similarity threshold of 1). These results indicate quite a high degree of overlap within the mobile space. These results suggest that there is a reasonable degree of repetition among mobile searches. Users are often submitting queries that are at least partially similar to queries that have been submitted previously. Recent work in the Web search space suggests that this degree of repetition can be exploited by boosting results from previous searches on similar queries when responding to a new target query [Smyth et al. 2004, 2005], and these results suggest that similar techniques may be used to good effect in mobile search.

5.5 Topics and Interests

So far we have examined how mobile subscribers search by analyzing the nature of their queries, how they formulate these queries as well as the patterns of query modifications. However, one important question remains: what types of information do mobile users search for? In this section, we will look at the various topics mobile subscribers are interested in by classifying the terms that they are using in their queries.

To carry out the analysis, we Classified a subset of the queries into a variety of different categories. We extracted the top-500 queries (in terms of their frequency) and manually assigned each to one of the 16 classes shown in Table VII[4].

[4]These categories were generated explicitly from our dataset. Initially we were hoping to use the same category sets outlined in Jansen et al. [2000] and Spink et al. [2001]. However we found that our data did not fit well into these Web-specific categories, thus we generated our own category list.

Table VII. List of Query Categories

No	Category	No	Category
1	Adult	9	Shopping & eCommerce
2	Multimedia	10	Mobile Applications, Websites & Technologies
3	Email, Messaging & Chat	11	Sport
4	Search & Finding Things	12	Auto
5	Entertainment	13	News & Weather
6	Games	14	Local Services
7	Unknown/Unclassified	15	Information
8	Socializing & Dating	16	Employment

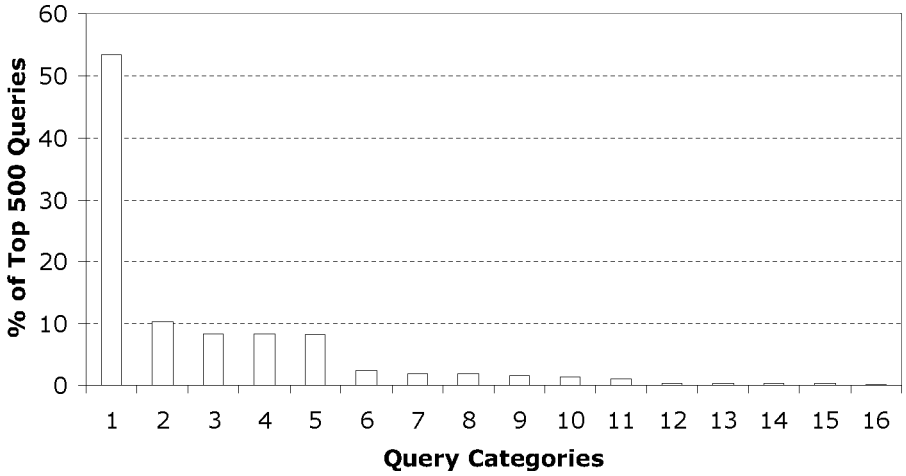

Fig. 11. Distribution of the top-500 queries across the query categories (total number of queries = 97894).

Figure 11 shows the distribution of the top-500 queries across the various query categories. The most popular type of query was *adult-related*, for example, "free porn", "porn", "sex" and "sex stories", which accounted for 53% of the top-500 queries. These results are consistent with Kamvar and Baluja [2006] who also found adult-related queries to be the most popular category of queries with users of Googles XHTML search. We also see clear parallels with early Web search studies, most of which found adult-related searches to be popular among early Web users [Jansen et al. 2000; Spink et al. 2001; Silverstein et al. 1999].

Next in popularity was *Multimedia* queries with 10% of the top-500 queries falling into this category. These included queries like "ringtones", "free wallpapers", "mp3 downloads" and "themes". The *Email*, *Messaging* and *Chat* category was also quite popular with our users. Approximately 8% of the top 500 queries are associated with this category. Sample queries from this category include "hotmail", "chat" and "msn messenger".

The Search category included queries for other search engines and search services like "aol", "tagtag" and "msn". These accounted for 8% of searches. Entertainment queries which also constituted about 8% of the top-500 queries include queries like "cinema" and "tv listings" as well as celebrity/personality searches including "britney spears" and "paris hilton".

The relative popularity of our query categories share similarities with the existing Google mobile search study [Kamvar and Baluja 2006]. Aside from adult content, searches for information on *Entertainment* and *Internet* and *Telecoms* were commonplace; note that in the Google study, the *Internet and Telecoms* category includes some of the queries we have classified as *Multimedia* in our analysis, for instance, "free ringtones".

6. DISCUSSION

We have analyzed the logs of a major European mobile operator. The data consisted of over 30-million mobile Internet requests generated by more than 600,000 unique European mobile subscribers over a 24-hour period in late 2005. Our goal was to identify key behaviors in mobile browsing and mobile search focusing on: (1) the usage differences between search and browsing behaviors, (2) the relationship between handsets and the preferred mode of information access, and (3) the emergence of mobile search behavior and in particular the early indicators with respect to how users are actively searching for information on their mobile devices.

This study is unique in a number of important respects. For a start, it investigates the use of both browsing and searching on a range of different mobile devices. In addition, it covers the full gamut of mobile content by considering XHTML and WML-based content. From a search perspective, it covers more than 30 different mobile search engines including the major players such as Google and Yahoo in addition to the wide range of niche mobile search engines that are in existence (e.g., Click4Wap and TagTag). Finally, unlike recent mobile search studies [Kamvar and Baluja 2006], this work provides an in depth analysis of mobile search behaviors that goes beyond the high-level usage metrics (e.g., query size) to consider important issues such as a detailed analysis of query modification strategies.

In this section, we summarize and discuss our key findings and their implications for the future under three important headings: (1) a comparison of browsing versus search activities; (2) understanding the usage of devices and search engines; and (3) an analysis of emerging search behavior.

6.1 Browsing vs. Search

To begin with, we mentioned at the start of this article how the mobile Internet world continues to be dominated by operator portals as the primary sources of mobile content. It was perhaps unsurprising, therefore, to find that approximately 94% of all mobile Internet sessions were classified as browsing-only sessions, that is, these sessions included no interactions with mobile search engines. Thus, only 6% of sessions (approximately 73,000 sessions) contained search engine interactions. We called these search sessions, and it was of course noted that these sessions were typically comprised of a mixture of search and browsing activities. Moreover, we found that 8% of subscribers engaged in at least one search session within the data analyzed.

On the basis of these data, it is not unreasonable to conclude that search activity represents a relatively small fraction of mobile Internet activity. However, when we compared the activity levels of search users (those users who did

engage in search sessions) to browsing users (those users who only browsed for information), we found a significant difference in terms of their duration online, bytes downloaded, and their average number of requests. By these metrics, a typical search session is approximately 3-4 times richer than a typical browsing session. Moreover, we also found that the average search user engaged in 2.8 online sessions per day compared to only 1.8 sessions per day for a browsing user.

As mentioned previously, there are two ways to read these results, and the correct interpretation depends greatly on the direction of the arrow of causality. One interpretation is that mobile search engines are an important activity driver on the mobile Internet and that by encouraging users to avail themselves of mobile search services, operators will see a significant growth in online activity. Another interpretation is that search engines are being used by early adopters who are more active mobile citizens by their very nature and that it is this inherent interest in trying out new features that is driving usage increases for this relatively small proportion of users.

In all probability it is likely that a significant proportion of the search users do conform to the early adopter stereotype. However, that being said, this does not detract from the additional online activity that their search behavior seems to generate. For example, it is worth pointing out that of the 2.8 daily sessions, referred to previously, generated by search users, approximately half of these are actual search sessions and the other half are browsing sessions. Therefore, a search user is engaging in a similar amount of browsing sessions as a regular browsing user but, in addition, is generating an additional set of search sessions. This suggests that search users are more actively engaged in using the mobile Internet in the search for new information since search plays an important role in content discovery. These search users clearly have information needs that they feel are best addressed by the use of mobile search engines rather than operator portals, and their use of these mobile search services leads to a richer online experience, benefiting the user and operator alike.

6.2 Search Engines and Devices

Our results indicate that the mobile search space is more fragmented than the traditional world of Web search. Although Google appears to dominate mobile search (attracting 76% of mobile searches), there remains a relatively long tail of competing search engines compared to traditional Web search. The remaining 24% of mobile searches are answered by up to 30 different mobile search engines, many of which offer niche search services dedicated to the mobile Internet. It is also worth noting that the dominance of Google is at least assisted by the emergence of Google-enabled handsets (e.g., some Sony-Ericsson handsets) that provide for one-click access to Google.

As expected, the majority of mobile handsets used for MI activities (browsing and searching) were found to be modern, feature-rich handsets combining high-resolution color screens with advanced text entry and XHTML support. When we compared device usage between browsing and search activities, we found a tendency towards more sophisticated handsets among the search users.

In the main, this meant a greater number of devices with higher resolution screens and WAP 2.0 support, another indication that search users are more likely to be tech-savvy early adopters.

Finally, we were surprised to see little variation in the text-input features offered by the popular browsing and search handsets. For example, none of the top-10 handsets used in either browsing or search included extended keyboards despite the fact that search includes an increased need for text-entry support. One reason for this is that such handsets remain relatively expensive and rare among the general population of mobile users and even the early adopters probably do not prioritize such mobile information access features when making their purchase decision, at least not ahead of more popular features such as cameras or MP3 capabilities. In recent time, more and more handsets have been launched with advanced text-entry features, such as extended keyboards, and so it is likely that, in the future, a greater proportion of MI handsets will include search facilities.

6.3 Emerging Mobile Search Behavior

The primary motivation for our study was to improve our understanding of mobile search behavior as opposed to alternative information access behaviors such as browsing. To this end, significant attention had been devoted to a fine-grained analysis of mobile search behaviors by adopting a similar methodology to that used by previous Web search studies (see Silverstein et al. [1999], Jansen et al. [2000], and Spink et al. [2001]). Based on our analysis of over 400,000 search queries by almost 50,000 mobile users, we have identified a number of key findings relating to mobile search behavior. Moreover, during the course of this work, we have compared these findings to previous Web search studies [Silverstein et al. 1999; Jansen et al. 2000; Spink et al. 2001] as well as the only other mobile search study in existence to date [Kamvar and Baluja 2006]. This analysis and comparison has highlighted a number of important similarities and differences between mobile and Web search which we will now discuss.

When one considers mobile search, it is reasonable to assume that the limited text-input capabilities are likely to have an impact on the way that we search for information on the MI compared to the traditional Web. Certainly we have found that the use of advanced search features is all but absent from mobile search; less than 3% of searches use any form of advanced search features compared to approximately 10% in Web search. One might also reasonably expect shorter queries in mobile search, although this turns out to be more complex than it first appears. To begin with, we found that mobile search queries are slightly shorter than their Web counterparts. Mobile queries average 2.06 terms per query overall compared to approximately 2.35 terms per query for early Web search usage [Silverstein et al. 1999], although recent results suggest that Web queries have now reached 3 terms per query [ZDNet 2006b]. Interestingly, we find a difference between query lengths when we compare Google mobile searches to non-Google mobile searches. For example, the average query length for non-Google mobile searches is just 1.53 terms per query compared to 2.22 terms per query for Google mobile searches (in line with the results reported

by Kamvar and Baluja [2006]). These lower query lengths are perhaps not so surprising, since all mobile devices are not well adapted for text input. However, the discrepancy between Google mobile and non-Google mobile searches is an interesting one. It suggests that perhaps mobile searchers who use Google are searching the mobile Internet using similar queries (at least in terms of term count) to those they might use on the traditional Web where as those users that have come to adopt dedicated mobile search services have also come to adapt their search patterns for the MI.

It is also interesting to note that we found a much lower incidence of unique queries among mobile queries (22% of mobile queries are unique) compared to Web search queries (approximately 57%–62% queries are unique). In addition, we found a significant degree of query overlap, for example, 60% of queries shared 50% of their terms with other queries. This indicates that at the present time mobile searchers have a more limited vocabulary, suggesting that they are searching for a more limited variety of online content than their Web counterparts. There are a number of factors that are likely to contribute to this. At present mobile Internet content is more limited in scope than general Web content. For example, most mobile users (at least initially) are exposed to mobile content through the operators portal. The content on most of these portals falls into fixed categories such as news, weather, or sport and is therefore quite restricted. Users who want alternative content must look beyond the walled-garden of the operator's portal.

Whether these smaller vocabularies are likely to expand and evolve is unknown at this point, but it seems reasonable to expect that as more content becomes available on the greater MI, user interests will broaden and their search vocabulary will evolve. Certainly we have seen similar patterns of change in the Web at large over the past few years with respect to query lengths, the type of queries submitted, and the topics of interest. For example, in the Excite 1997a study [Jansen et al. 2000], adult content represented 16.8% of their query sample, while a subsequent analysis of Excite data in 2001 [Spink et al. 2002] demonstrated a shift away from adult content. The percentage of adult content within the query sample dropped by almost half to 8.5%. It is therefore reasonable to assume that a similar effect will occur on the mobile Internet especially as off-portal search and browsing gathers pace. Indeed, this is already evident from our results. For example, as already stated, Google's mobile queries (which mainly relate to searches for off-portal content) are longer (2.22 terms) than the queries submitted to other mobile search engines (1.53 terms), many of which represent searches for on-portal content. Moreover, for the 40,341 unique search query terms in our dataset, 32,686 of these were associated with Google queries. This means that over 80% of the unique terms used in all queries were specific to Google, indicating a broader vocabulary than some of the other search engines we analyzed.

Finally, while we have found strong parallels between current mobile search behavior and early Web searchers (the prevalence of short, vague queries and similar interests such as adult content), we also found a number of interesting differences which suggest that mobile searchers may be finding it more difficult to locate relevant information using search. Mobile searchers engage in more

searches per session than their Web counterparts. For example, we found the average mobile searcher engages in 5.8 searches per session compared to only 4.9 (Excite 1997b) for comparable Web searchers. Also, although we found similar degrees of query modification in both mobile and general Web search (23% for mobile search and 22% for the Excite 1997a dataset), mobile search is characterized by a much lower proportion of initial queries (17% for mobile search compared to 35% and 48% for the Excite studies) and a higher proportion of identical/repeat queries (58% for mobile search).

7. IMPLICATIONS

One of the motivations for this work has been to benchmark the early stages of mobile search with a view to tracking its evolution going forward. We believe that this is especially important if we are to understand the similarities and differences between mobile and Web search. Of course one of the key reasons for this type of work is to better understand the nature of mobile search with a view to improving future generations of mobile search engines. With this in mind, there are a number of opportunities for improvement which stem from the results that we have highlighted which we will briefly outline in the following sections.

7.1 Exploiting Query Repetition

We have already mentioned the limited query vocabulary and significant overlap between queries in mobile search, highlighting the high degree of query repetition that naturally exists among the search terms used by mobile searchers. This is important because it suggests a certain degree of regularity among the search patterns of mobile users, a regularity that can be exploited in order to improve search precision.

Specifically, the work of [Smyth et al. 2004, 2005] demonstrates how such query repetition can be harnessed by reusing the past result selections of searchers in response to similar queries. By promoting such selections within a result list, it was possible to significantly improve result relevance. In particular the potential for improvement was found to be closely correlated to the degree of overlap and repetition between queries, which suggests that similar improvements may be available to future mobile search engines.

7.2 Supporting Query Modification

The frequency of query modification noted in mobile search represents another opportunity for improving mobile search services. We found that approximately 23% of search queries were modifications of previous queries. This suggests that users did not find what they were looking for initially and chose to modify their query, typically by adding, deleting, or replacing query terms to produce a more precise query. It also suggests a very significant opportunity in mobile search to assist users by suggesting or recommending useful related queries. Indeed the area of query recommendation and refinement has received considerable interest within the information retrieval community over the past few years [Zhang and Nasraoui 2006; Baeza-Yates et al. 2004] which could now be considered in

the light of mobile search services and interfaces. Particularly relevant in this context, especially given the query repetition rates found for mobile search, is the query recommendation work presented by Balfe and Smyth [2004]. Specifically, this approach to query recommendation exploits past search behaviors and query repetition for query recommendation, facilitating the recommendation of related queries at the search result-level and ranking these queries according to how likely they are to match the user's information needs.

7.3 Enriching Mobile Content

Another interesting feature of the mobile Internet concerns the average size of MI pages. We found that approximately 52% of mobile requests within our dataset were for WML content. When we examined the size of these WML pages, we found that the average size was only 1036 bytes (about 1KB). In contrast, the average Web page is many times this size (typically 10–20KB). When we consider this from a search perspective, it highlights an important challenge facing mobile search engines. Put simply, search engines index page content by the terms that appear in the page. Fewer terms means a more limited indexing vocabulary, which in turn limits retrieval opportunities. It is well known that searchers often use query terms that do not appear in the pages that they are looking for and so these pages may not be retrieved. This problem is exacerbated by the limited content on MI pages and further by the limited query vocabulary of mobile searchers.

One potential solution to this problem involves the automatic enrichment of mobile page content by adding additional relevant terms for the purpose of indexing. For example, the work of Church and Smyth [2007] describes one such approach which uses existing search engines as a source of enrichment terms by converting a mobile page into a set of representative queries and then extracting popular terms from the snippet texts of the results that are returned by a search engine (such as Google) for these queries. By augmenting the mobile search index with these terms, it was possible to improve mobile search performance and increase the likelihood of retrieving a relevant page for even a vague mobile query.

8. CONCLUSIONS

In this article, we set out to investigate the information access behavior of more than 600,000 European mobile Internet (MI) users with a view to better understanding their access patterns and information usage strategies. We were especially interested in the growing usage of mobile search as a powerful off-portal content discovery technique and how mobile users were beginning to avail themselves of new mobile search services.

Overall we found mobile information access to be dominated by portal browsing activities, with 94% of sessions consisting of just browsing. That said, we found that 8% of mobile subscribers actively used mobile search sessions and also found that, as a general rule, these subscribers tended to be involved in much richer online behavior than their browsing-only counterparts in terms of their time online, bytes downloaded, requests, and sessions per day. In the

main, we found that MI users tend to avail themselves of the most modern types of mobile handsets with search users tending to use more sophisticated devices (especially in terms of screen size) than browsing users.

A key focus of this study in on mobile search, and, to that end, we have adopted the standard methodology used by Jansen et al. [2000] and Spink et al. [2001] to perform a detailed analysis of the behavior of mobile searchers. Our findings suggest that mobile search is in an early stage of maturity that is comparable to Web search during the mid 1990's. For example, mobile searchers appear to avail themselves of a wide range of different search engines and, although Google dominates (with 76%) market share, approximately 30 search engines make up the remaining 24%. In addition, we found mobile search queries to be somewhat shorter than their Web search counterparts—2.06 vs. 2.3 query terms for mobile and Web search, respectively—and minimal use of advanced search features. Interestingly, we found a relatively high rate of query modification among mobile searchers (23% of queries were modifications of previous queries) with almost half of the modifications involving a substitution of query terms, a significantly higher substitution rate than that found in Web search. Finally, we also observed how mobile search tended to use a more limited vocabulary than Web searchers which is more than likely influenced by the relatively limited mobile content that is attracting niche user interests, for example, 50% of the top-500 mobile search queries refer to adult content.

The analysis so far has suggested a number of avenues for future research that may lead to potential improvements in mobile search services. For example, high query repetition and modification rates suggest significant opportunities to harness prior search behaviors in order to improve future searches for related queries. In addition, content enrichment techniques may be one way to enhance limited mobile content for improved indexing.

Of course, as with any such study, there are caveats and limitations. For example, our research has been confined to European mobile subscribers and excludes important markets such as the US and Asia. That said, the US-centric study by Kamvar and Baluja [2006] complements our own work. In common with other search studies, it has not been possible to capture information about the behavior of users beyond the search results page, and so it is not possible to confidently report on important issues such as how frequently searchers find what they are looking for. In addition, this information would be especially useful when it comes to better understanding as to why mobile search sessions appear to be so much richer than simple browsing sessions as discussed previously. Finally, our study provides an analysis into mobile information access at a particular point in time. As the mobile Internet and mobile search gathers pace, we need to produce subsequent analyses that describe the changes in behavior, just as the various Web search studies have done in the past. Thus this study is viewed as one of a series of mobile search studies that will help to track the evolution of mobile search over time.

In closing then, there are many who believe that the mobile Internet is a sleeping giant that will soon awaken to become the dominant form of anytime-anywhere information access. Certainly the advent of next-generation handsets

and networks have helped to significantly improve the quality of the online experience for mobile users, and today a significant percentage of users use their mobile phones to access online information and services. Our analysis has confirmed that, unlike the Web world, mobile information access continues to be dominated by portals and browsing rather than search engines and query-based search. That said, there is evidence to suggest that the role of mobile search is on the increase with many of the heavyweights of Web search now looking to the mobile sector as a new frontier. The early adopters of mobile search are making a significant usage impact already but continue to struggle with the peculiar challenges presented by the mobile space, especially in relation to limitations of screen space and text input. We believe that these challenges can only be overcome by taking a fresh look at mobile search, one that recognises the needs of the individual as a key element in the next generation of mobile search engines that are able to adapt to the personal preferences of searchers.

REFERENCES

BAEZA-YATES, R., HURTADO, C., AND MENDOZA, M. 2004. Query Recommendation using query logs in search engines. In *International Workshop on Clustering Information Over the Web (ClustWeb)*.

BALFE, E. AND SMYTH, B. 2004. Collaborative query recommendation for Web search. In *Proceedings of 16th European Conference on Artificial Intelligence* (ECAI'04). IOS Press, 268–272.

BALFE, E. AND SMYTH, B. 2005. An analysis of query similarity in collaborative Web search. In *Proceedings of the 27th European Conference on Information Retrieval* (ECIR'05). 330–344.

CELLULAR NEWS. 2006. 2.5 billion mobile phones in use. http://www.cellular-news.com/story/19223.php. (Accessed March 2007).

CHURCH, K. AND SMYTH, B. 2007. Mobile content enrichment. In *Proceedings of the International Conference on Intelligent User Interfaces* (IUI'07). ACM Press, 112–121.

HÖSCHER, C. AND STRUBE, G. 2000. Web search behavior of Internet experts and newbies. In *Proceedings of the 9th International World Wide Web Conference on Computer Networks. Int. J. Comput. Telecomm. Netw.* North-Holland Publishing Co., 337–346.

IPOS INSIGHT. 2006. Mobile phones could soon rival the PC as worlds dominant Internet platform. http://www.ipsos-na.com/news/pressrelease.cfm?id=3049. (Accessed March 2007).

JANSEN, B. J. AND SPINK, A. 2005. An analysis of Web searching by European AlltheWeb.com users. *Inform. Process. Manag. 41*, 2, 361–381.

JANSEN, B. J., SPINK, A., AND PEDERSEN, J. 2005. A temporal comparison of AltaVista Web searching: Research articles. *J. Amer. Soci. Inform. Scien. Technol. 56*, 6, 559–570.

JANSEN, B. J., SPINK, A., AND SARACEVIC, T. 2000. Real life, real users, and real needs: A study and analysis of user queries on the Web. *Inform. Process. Manag. 36*, 2, 207–227.

KAMVAR, M. AND BALUJA, S. 2006. A large scale study of wireless search behavior: Google mobile search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'06). ACM Press, 701–709.

OPEN MOBILE ALLIANCE. 2002. WAP wireless markup language specification (WML). http://xml.coverpages.org/wap-wml.html. (Accessed March 2007).

SILVERSTEIN, C., MARAIS, H., HENZINGER, M., AND MORICZ, M. 1999. Analysis of a very large Web search engine query log. *SIGIR Forum 33*, 1, 6–12.

SMYTH, B. 2002. The plight of the mobile navigator. *MobileMetrix*.

SMYTH, B., BALFE, E., BOYDELL, O., BRADLEY, K., BRIGGS, P., COYLE, M., AND FREYNE, J. 2005. A live-user evaluation of collaborative Web search. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence* (IJCAI'05). 1419–1424.

SMYTH, B., BALFE, E., FREYNE, J., BRIGGS, P., COYLE, M., AND BOYDELL, O. 2004. Exploiting query repetition and regularity in an adaptive community-based Web search engine. *User Model. User Adapt. Interact. 14*, 5, 383–423.

SMYTH, B. AND COTTER, P. 2002a. Personalized adaptive navigation for mobile portals. In *Proceedings of the 15th Eureopean Conference on Artificial Intelligence* (ECAI'02). IOS Press, 608–612.

SMYTH, B. AND COTTER, P. 2002b. The plight of the navigator: Solving the navigation problem for wireless portals. In *Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems* (AH'02). Springer-Verlag, Berlin, Germany, 328–337.

SPINK, A., JANSEN, B. J., WOLFRAM, D., AND SARACEVIC, T. 2002. From e-sex to e-commerce: Web search changes. *IEEE Comput. 35*, 3, 107–109.

SPINK, A., WOLFRAM, D., JANSEN, M. B. J., AND SARACEVIC, T. 2001. Searching the Web: The public and their queries. *J. Amer. Soc. Inform. Sci. Technol. 52*, 3, 226–234.

VAN VEEN, N., DE LUSSANET, M., AND MENKE, L. 2005. European mobile forecast: 2005 To 2010. Forester Research.

WOLFRAM, D., SPINK, A., JANSEN, B. J., AND SARACEVIC, T. 2001. Vox Populi: The public searching of the Web. *J. Amer. Soc. Inform. Sci. Technol. 52*, 12, 1073–1074.

ZDNET. 2006a. April 2006 search engine market shares. http://blogs.zdnet.com/ITFacts/index.php?p=10999. (Accessed March 2007).

ZDNET. 2006b. Yahoo: Searches more sophisticated and specific. http://blogs.zdnet.com/micromarkets/index.php?p=27. (Accessed March 2007).

ZHANG, Z. AND NASRAOUI, O. 2006. Mining search engine query logs for query recommendation. In *Proceedings of the 15th International Conference on World Wide Web* (WWW'06). ACM Press, New York, NY. 1039–1040.

Received August 2006; revised January 2007; accepted February 2007