Author(s):	Heimonen, Tomi
Title:	How Do Users Search the Mobile Web with a Clustering Interface? : A Longitudinal Study
Year:	2012
Journal Title:	International Journal of Mobile Human Computer Interaction
Vol and number:	4 : 3
Pages:	44-66
ISSN:	1942-390X
Discipline:	Computer and information sciences
School /Other Unit:	School of Information Sciences
Item Type:	Journal Article
Language:	en
DOI:	http://dx.doi.org/10.4018/jmhci.2012070103
URN:	URN:NBN:fi:uta-201403271272

International Journal of Mobile Human Computer Interaction

July-September 2012, Vol. 4, No. 3

Table of Contents

Editorial Preface

Research Articles

How Do Users Search the Mobile Web with a Clustering Interface?
A Longitudinal Study

Tomi Heimonen, University of Tampere, Finland

ABSTRACT

Category-based search result organization holds promise as a means of facilitating mobile information access. This paper presents the results of a longitudinal user study that investigated how a mobile clustering interface is used to search the Web. The author describes the participants' search behavior and discusses the benefits and limitations of category-based result access. Study results show that category-based interaction was considered situationally useful, for example when the participants had problems describing their information need or needed to retrieve a subset of results. The paper proposes design guidelines for category-based mobile search interfaces. These include improved strategies for presenting the categories in the search interface, the need to improve the categorization methods to provide more representative category structures, and accounting for the contextual aspects of mobile information needs.

Keywords: *Design Guidelines, Longitudinal Evaluation, Mobile Search, Search Interface Design, Search Result Clustering*

INTRODUCTION

Mobile Internet has become an indispensable medium of information access for users around the world. It is used increasingly as a tool for communication, information gathering, transactions and establishing one's online presence (Taylor et al., 2008). A recent survey by Kaikkonen (2011) highlights the growing importance of search as a mobile information access method. The use of mobile Web search has increased significantly between 2007 and 2010, with tasks such as searching for contact information, locations and routes being even more common than on the desktop. Mobile search has been identified as an important information access tool in various social situations; specifically as a means to satisfy information needs as they arise (Church & Oliver, 2011; Heimonen, 2009). How people look for and interact with information on mobile devices is triggered by different contextual needs such as time, location and activity (Church & Smyth, 2009; Hinze, Chang, & Nichols, 2010; Sohn, Li, Griswold, & Hollan, 2008), and should therefore inform search systems design.

Commercial search engine providers offer mobile-oriented search services to facilitate mobile information access. Although these services are specifically designed for mobile devices and make use of many useful features such as location sensing and voice interaction, ultimately the search results themselves are of-

DOI: 10.4018/jmhci.2012070103

ten presented in flat, ranked result lists. Previous studies on mobile Web search patterns suggest that users experience difficulties with these interfaces (Kamvar & Baluja, 2006; Church, Smyth, Bradley, & Cotter, 2008). The search data analyzed by Kamvar and Baluja (2006) shows that mobile searchers did not explore the results actively, as only about 9% of searches went beyond the first result page. Results fulfilling the users' information need may remain unseen simply because of ambiguous queries that do not produce relevant results in the first result page. Mobile searchers also remained focused on their initial search topics; only 25% of the subsequent queries were not directly related to the first. This lack of exploration can be explained by the relatively higher cost of interactions in the mobile environment, e.g., slow loading times and overhead of browsing through result pages. The study by Church et al. (2008) highlights other problems with mobile search user experience: almost 90% of queries and nearly 60% of search sessions do not lead to any result selections by the user. It is likely that in some cases the conventional approach to mobile search adopted by search engines results in users failing to find relevant information with the result lists (Church et al., 2008).

One of the key issues is that the ranked result list does not provide an effective overview of the themes present in the result set. This makes search challenging when one's needs go beyond simple keyword lookup, or when the query is hard to specify. Users engaging in more exploratory search of an unfamiliar topic may require additional help in understanding the terminology and structure of the result set (White, Drucker, Kules, & Schraefel, 2006). Such assistance can be provided by categories, which can help information seekers make sense of search results and decide which actions to pursue (Hearst, 2006). Category-based search and browsing is commonly used in online stores such as Amazon.com to access titles organized into consistent product hierarchies. Previous research on mobile search interfaces suggests that categories could also be helpful in mobile search situations (Carpineto, Mizzarro, Romano, & Snidero, 2009; De Luca & Nürnberger, 2005; Heimonen & Käki, 2007; Karlson et al., 2006; Machado et al., 2009).

So far the benefits of category-based mobile search interfaces have been demonstrated in controlled laboratory studies, which motivated us to investigate category use in a realistic context of use in a four-week user study. We addressed the following research questions:

1. How are the categories used for result access during naturalistic mobile Web search?
2. In what kind of search scenarios do the categories benefit the users, how, and what are their limitations?

Our findings describe the participants' search behavior with a clustering search interface and relate it to previous studies of mobile search behavior and cluster use in desktop and mobile search interfaces. We also identify benefits and limitations of category-based search grounded on the findings of this study and those of previous research. We propose design suggestions to alleviate these issues. These contributions are likely to be of interest to audiences in the mobile information access community.

The remainder of the article is organized as follows. First, we review previous research on category-based search interfaces and mobile search interface evaluation methodology. Following, we describe the interface utilized in the study, the details of the longitudinal study and its results. The paper concludes with a discussion of the key results and presents the design implications inspired by these insights.

RELATED WORK

Methods for Organizing Search Results

The main methods for organizing search results into category structures are category systems and document clustering (Hearst, 2009). In category systems, documents are assigned

into predefined category hierarchies either manually or automatically by using classification algorithms (e.g., Chen & Dumais, 2000; Kules & Shneiderman, 2008). The benefit of category systems is that the resulting structures are typically logical, consistent and familiar to users (Hearst, 2009). One type of classification employed for searching and browsing Web site content is faceted classification. Each document is classified along multiple dimensions called facets, which can be used to for example browse image collections by theme, artist or location (Yee, Swearingen, Li, & Hearst, 2003). Although the usefulness of classification in practical Web search scenarios is limited by the need for some manual assignment and classification construction, Hearst (2009) notes that category systems have been found to be more usable than ranked result lists or clustering in previous research.

Clustering techniques form collections of documents based on their similarity. Clustering can consider the overall similarity of the documents (Cutting, Karger, Pedersen, & Tukey, 1992), or be based on shared features such as frequent words and phrases (Käki, 2005a). Clusters provide an overview of the topic with terms to inform query reformulation, and results related to the same subtopic are presented together, which allows fast retrieval of items of interest (Carpineto et al., 2009a). Clustering can be helpful for ambiguous, vague or broad informational queries, which can be problematic for traditional search interfaces (Carpineto, Osiński, Romano, & Weiss, 2009). For example, the ambiguous query "jaguar" could result in clusters "big cat," "operating system," and "car," reflecting the different subtopics among the results. The usage patterns identified in mobile search, such as short queries, low click-thru and lack of results exploration, make clustering appealing for mobile search (Carpineto et al., 2009a). Unlike classification construction, clustering can be fully automated and the algorithms are applicable even for short documents such as Web search results (Hearst, 2009). In contrast to classification, the drawbacks of clustering include the lack of predictability and consistency of the outcomes, and challenges in providing understandable labels (Hearst, 2006). Carpineto et al. (2009b) advocate the use of *description-centric* clustering algorithms that are designed specifically for clustering search results and account for both clustering and label quality. These algorithms aim at ensuring the comprehensibility and conciseness of the labels, in addition to establishing the relationship between the label and the cluster's documents clearly. The clustering algorithms utilized in our interface are such description-centric approaches, as they base the clustering on the label extraction process, and attempt to provide understandable labels by using common terms extracted from search result captions.

Category-Based Interfaces for Mobile Web Search

Clustering Interfaces

Several clustering search engines and interfaces have been proposed for mobile Web search. Carpineto, Della Pietra, Mizzarro, and Romano (2006) introduced Credino, a clustering search engine for mobile devices based on concept lattices, a form of hierarchical clustering. In their approach the categories are arranged as an expanding hierarchy where the cluster labels act as links to result pages. Their small-scale user study demonstrated that search result clustering provides higher performance than ranked result lists. In a follow-up study, Carpineto et al. (2009a) compared desktop, PDA and mobile phone search interfaces (both with and without clustering) in a 72-participant between subjects experiment over four tasks that represented different information needs (e.g., informational, transactional and navigational). Their results show that clustering is more effective than the ranked result list in terms of search performance on PDAs and mobile phones. In a closer examination of individual tasks, they identified benefits and drawbacks of clustering. For example, clustering failed to provide relevant category labels in the fourth task, leading

to low performance. On the other hand, in the first task clustering performed well because the cluster hierarchy provided good hints about an otherwise unfamiliar search topic and helped narrow the search down to the correct result. They also noted the need for collecting more evidence about clustering performance with more diversified tasks specific to real mobile search scenarios.

In our previous work we presented Mobile Findex, a mobile Web search interface that provides a flat list of clusters computed on the basis of the most common words and phrases within the search result captions (Heimonen & Käki, 2007). Our 16-participant laboratory user study compared the ranked result list to the clustering interface over 12 information-seeking tasks with topics such as trivia and shopping. The results suggest that the clustering interface could offer search performance comparable to a traditional mobile Web search interface, with participants preferring Mobile Findex because of its perceived efficiency, suitability for the type of tasks used in the study, and ease of finding results. The overview and filtering capabilities provided by the clusters were judged to be more essential to the user experience than pure search performance. Further analysis of the results showed that clustering aided in ambiguous queries because the cluster labels could be used for disambiguation and drilling into relevant sets of results (Heimonen, 2008).

Machado et al. (2009) suggested a clustering search interface with three distinct display modes: (1) cluster list that displays the first result of each cluster, (2) classic cluster label list, with the cluster label displayed alongside the other metadata related to each result, and (3) full-screen display. Although a systematic evaluation is not reported, the authors note concerns about how the clustering interface may interfere with users' established search habits.

Classification and Faceted Browsing Interfaces

In addition to clustering interfaces, also classification and faceted search interfaces have been suggested for mobile search. Buchanan, Jones, and Marsden (2002) introduced LibTwig, a category-based overview interface for mobile digital libraries. The LibTwig user interface organizes results as an expanding outline tree, which the user can explore by selecting tree nodes until the actual result documents are reached. Evaluations of LibTwig suggest that non-expert Web users prefer the outline approach because it provides them with a good overview of the result set. De Luca and Nürnberger (2005) proposed an approach in which they use several classification methods, such as semantic classification of the query terms and the use of bookmarks to provide disambiguation information for each result. Karlson et al. (2006) introduced FaThumb, a mobile search interface based on faceted hierarchies and browsing. Their evaluation showed that faceted browsing is faster than keyword search for less specific information needs.

Other Methods for Facilitating Information Access in Mobile Search

Other alternate approaches to ranked result list for presenting search content have also been explored. Jones, Jones, and Deo (2004) proposed the use of keyphrases automatically generated from the result document content as an alternative form of result surrogate. The results of their study indicate that when there the title provided for a search result is missing or poor, keyphrases could aid the user in making sense of the result. Church, Smyth, and Keane (2006) suggested the use of keywords from related queries as a more economic alternative to snippets. Their results indicate that related queries are a viable alternative in mobile result presentation, especially as they provide a balance between informativeness and screen space usage. Other approaches have focused on utilizing other users' queries in an interactive manner. Arter, Buchanan, Jones, and Harper (2007) carried out a study with the Questions-not-Answers (QnA) interface, which shows location specific queries from other users in a

map-based interface. One of their key takeaways is that location-based search interfaces would benefit from making the distinction between functional information about a location and less familiar, playful information. Their results also show that the insights and sense of place that can be gained and reflected by queries vary by location type. Church, Neumann, Cherubini, and Oliver (2010) introduced SocialSearchBrowser (SSB), which is similar to the QnA interface but also provides filtering, search and social networking features. Their results showed that the participants were not overly concerned about sharing their queries and answers to questions posed by other users. The majority of users also preferred human-generated content to traditional search engine results, especially for personal and time-sensitive information needs.

Evaluating Mobile Web Search Interfaces

A common theme to the previous research on category-based mobile search interfaces is the use of lab-based evaluations. Evaluating mobile interfaces in the lab restricts the breadth of possible research questions as contextually emerging information needs and the strategies used to address them cannot be studied reliably. Evaluating the exploratory features and exploratory use of an interface adds its own complexity because exploration is not adequately captured by typical precision-recall metrics (White, Muresan, & Marchionini, 2006).

One promising approach for studying search systems *in situ* is longitudinal evaluation, in which participants use the system as a part of their every day information seeking over an extended period of time (e.g., Käki, 2005a; Wilson & Schraefel, 2008). Longitudinal studies can be especially useful for evaluating search user interfaces because they allow for observing changes in the use of the system over time and over a range of varying information needs (Hearst, 2009).

Previous longitudinal studies of mobile search interfaces have been reported in literature. In the QnA study (Arter et al., 2007), people took part in the study over a period of four and a half days. Usage logs and diary entries were combined with telephone interviews to canvass initial impressions and conduct post-study interviews. This approach was instrumental in uncovering unexpected behavioral patterns that would have been difficult if not impossible to discern in a short lab session. Amin, Townsend, Ossenbruggen, and Hardman (2009) investigated location-based mobile search behavior in a 12-day study that incorporated web-based diaries, search interaction logging and interviews. Their method enabled them to capture explicit search behavior (queries), intentions (motivation behind search) and the context of the search. Church et al. (2010) studied their SSB prototype with 16 participants for the duration of one week, followed by a post-study survey to gather subjective feedback. Our methodology is very similar to this last approach and was motivated by trading off increased data collection (such as daily diary entries and accurate context) with longer exposure to the search interface and a decreased burden of attendance for the participants.

MOBILE CLUSTERING WEB SEARCH INTERFACE

We designed and developed a mobile search interface prototype for the purposes of our longitudinal study. The Mobile Findex interface was implemented on top of the Findex search and clustering framework (Käki, 2005b), which was adapted to utilize Microsoft Live Search (http://www.bing.com/toolbox/bingdeveloper/) as the backend search engine. We incorporated design guidance from several sources, including the guidelines by Jones and Marsden (2006) and previous studies on category-based mobile search interfaces. Jones and Marsden (2006) suggest that the two main design goals for a mobile search interface should be to allow the users to quickly evaluate the relevance of the whole result set, and give the users enough information about individual results to make judgments on their usefulness. Given the benefits of clustering approaches for providing result overviews reported in previous research, using

clustering as the categorization method was a practical fit also for our prototype. We used previous systems such as Credino (Carpineto et al., 2009a) and the previous iteration of the Mobile Findex interface (Heimonen & Käki, 2007) to inform the interface design.

Effective interaction is paramount in mobile interfaces. Working within the constraints of available interaction methods (i.e., lack of touchscreens in the majority of phones at the time), our aim was to provide a straightforward and clear interaction model centered around two views: the category list and the result list. The design hypothesis was that separate views for overview and content reduce the need for vertical scrolling within a page while making better use of the limited display space. However, this design decision does force the user to adopt a "back-and-forth" oriented interaction style in accessing results between categories. A view containing both category labels and results would have necessitated a significant amount of up and down scrolling when switching between categories. With current higher resolution touchscreen devices an integrated design would be more effective as touch gestures allow for rapid scrolling and change of focus within the page.

Search Interface Layout

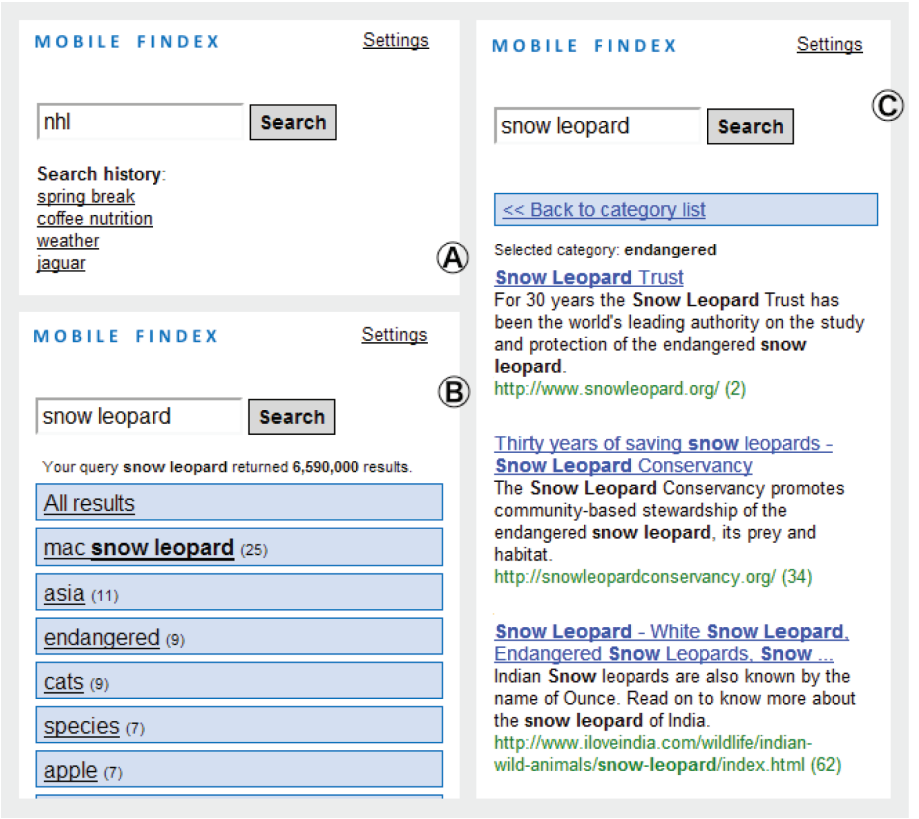
The search user interface of the prototype (Figure 1) consists of three views: the initial query view (a), the category list (b) and the result list (c). The query contains an input field for entering the query terms and the search history for the user, including up to four of the user's latest queries. The category view is used to present the search result clusters to the user. Each item in the category list consists of the category label and a number indicating how many results the category contains. The list is ordered based on the number of results contained in the cluster. The topmost item, 'All results,' can be used to access the full ranked list of results for the query. The result list presents the individual results in the ranking order of the underlying search engine, up to 15 results per page. Each result element is composed of the page title as a hyperlink to the result web page, the result caption and the URL. The number in parentheses after the URL indicates the ranking order of the result in the unmodified result list provided by the search engine.

Search Result Clustering Algorithms

The prototype incorporates two variations of a Web search result-clustering algorithm proposed in previous research. The statistical algorithm (Käki, 2005a) utilizes a simple frequency-based clustering technique that uses word and phrase frequencies in the search result snippets. It forms the clusters by maximizing the length of the cluster labels based on a predefined cluster limit. The resulting categories contain the results in which the terms (word or phrase) contained in the category label appear. The keyword context algorithm fKWIC (Käki, 2006) functions similarly to a keywords-in-context index where the occurrences of query terms are displayed together with the surrounding words. The overall functioning of the algorithm is similar to the statistical clustering algorithm. The key difference from the user's point of view is that with fKWIC, the query appears as a part of the cluster labels. A more in-depth discussion of the algorithm design is beyond the scope of this article and can be found elsewhere (Käki, 2005b).

The algorithms have some limitations common to other clustering approaches. The quality of the clusters, in terms of content and labeling, is dependent on the content of the underlying search result set. Since no semantic reasoning is applied the algorithms may potentially provide cluster labels that do not accurately describe cluster content. Keyword context categories can help with these problems to some extent since they integrate the users' query as a part of the label, providing a familiar context. As the clustering does not result in mutually exclusive separation of results, a result can exist under several clusters. This can be confusing for users who expect the clusters to cleanly separate results into orthogonal topics.

Figure 1. Clustering user interface **(A)** *Query view,* **(B)** *Category list, and* **(C)** *Result list*



Both algorithms performed the clustering using the first 150 results returned by the search engine. In our experience, this amount of results provides a good compromise between clustering quality and response time. The participants were provided with up to 10 categories per query. If the clustering algorithm could not compute a list of categories, the participant was shown the normal search result list instead. Participants could change both the clustering algorithm and desired number of clusters in the interface settings.

USER STUDY

The longitudinal user study of clustering interface use was arranged during September-November 2009. Prior to the full-scale study, an initial version of the prototype was pilot tested in a local media museum for a period of five weeks in October-November 2008. Changes were made to the log data collection based on the results to improve its robustness. The initial version of the prototype was further evaluated in a small-scale pilot study with eight participants in November-December 2008. The purpose of the pilot study was to gather feedback on the interface from active mobile Web users and verify the longitudinal study procedure. As a result, we made minor changes to the user interface and the subjective feedback protocol was expanded to allow for more in-depth data collection.

Participants

Seventeen active mobile Web and mobile phone users (five female, twelve male), aged from 18 to 54 years (mean = 28.1, SD = 8.6) were recruited for the study. Recruitment was undertaken using convenience sampling from local universities. The inclusion criteria for the study were experience with and active use of mobile phones, mobile Web and mobile search services. Participants were also required to carry a mobile broadband subscription or have daily access to Wi-Fi on their mobile phones.

All of the participants had over five years of experience using mobile phones, used a smartphone and had experience with mobile Web use with activity levels ranging from daily (13 participants) to weekly (4) use. Mobile Web search services were used daily by six participants, weekly by 10 participants and less frequently by one participant. The mobile Web search service of choice for all participants was Google Mobile Search. In addition to Google, five of the participants reported use of various other Web-based information seeking services such as maps applications, public transportation timetable search, other mobile Web search engines, and Web-based phone directory search.

The phones most frequently used by our participants were Nokia E and N series devices (10 participants) with keypad text entry and no touchscreen input, and with a screen resolution of 320 x 240 pixels. Six participants had touch-enabled devices including Apple iPhone 3G and iPod Touch (480 x 320 pixels), Nokia 5800 XpressMusic (640 x 360 pixels) and Samsung Jet (800 x 480 pixels). Participants with iPhone 3G and iPod Touch were provided with an interface adapted for their WebKit-based browsers.

Procedure

The participants were asked to use the Mobile Findex interface for the duration of four weeks. They were instructed to utilize the interface in their daily Web search activities "when it was feasible for them to do so". Following the guidelines by Shneiderman and Plaisant (2006), the participants were informed that they would not need to generate searches specifically for this study, but should utilize whichever information seeking tools they normally do, including other search engines. The participants received email notifications about the study twice a week.

The study consisted of three parts: an introductory interview, independent use of the prototype, and a post-study interview. The introductory interview lasted 30-45 minutes and consisted of the study protocol briefing and a discussion about the participants' previous mobile information access behavior. A subjective feedback questionnaire was administered via email after two weeks of interface use to gather initial experiences. The post-study interview lasted between 45 to 60 minutes and included questions related to the participants' mobile Web search behavior during the study. Background information and additional subjective feedback were collected with questionnaires during the introductory and post-study interviews. Four movie tickets with a total value of €32 were provided as compensation for participation.

Data Collection

The participants' interactions with the prototype were stored in server-side logs. For each search interaction, the following data were recorded:

- Time and date.
- Type of interface activity (e.g., executing a new search, selecting search result, etc.).
- Meta data related to search activity (e.g., query terms, selected category label, etc.).
- IP address of the device and user agent string of Web browser.

In addition to search interactions, quantitative subjective feedback was collected with questionnaires and interviews. User expectations and experiences were collected using the SUXES questionnaire tool (Turunen et al., 2009). Developed for the evaluation of interactive systems that include user-system dialogue (e.g., speech recognition), SUXES provides a way to study the differences between

the users' expectations and actual usage-based experiences. The gap between expectations and experiences highlights areas requiring the most improvement and ones showing the most promise from the users' perspective. The SUXES questionnaire was constructed of positively worded statements, such as "The search interface is useful." All responses were provided on a 7-point scale, ranging from complete disagreement (1) to complete agreement (7). The questionnaire contained nine statements each targeting one of the following aspects of the system: speed of use, pleasantness, clarity, error-free use, error-free functioning, ease of learning, intuitiveness, usefulness, and likelihood of future use. After receiving a brief introduction to the system, the expectation questionnaire was filled in by providing two values for each statement: the lowest acceptable level they would expect the system to function at, and the highest desirable level. Experiences that fall on these ends of the spectrum or beyond are considered problematic or successful features of the system.

The post-study interview included both open-ended and semi-structured questions regarding the clustering search interface. The topics included expectations and experiences of the interface, applicability of the clustering approach to one's own search strategies, perceived search success, and the frequency of clustering interface use during the study. In order to elicit responses in the post-study interview, we adopted the approach proposed by Benedek and Miner (2002), whereby the participants were asked to select words to describe their experience from two lists of positive and negative adjectives.

RESULTS

Analysis Methodology

The longitudinal study provided a considerable amount of quantitative and qualitative data. The search interaction log data were split into sessions and individual queries. A new session was identified when there was duration of three minutes between interactions with no identifiable link. For each session and query further quantitative measures were calculated. These include the number of query terms, session duration, and number of category and search result selections, as well as derived measures such as click-thru rate.

The subjective feedback contained both quantitative and qualitative elements. The quantitative data was analyzed using descriptive statistics and non-parametric statistical methods. The interview responses were analyzed using thematic qualitative analysis to identify themes related to the use and utility of the interface and clustering categories. These findings were related back to the quantitative analysis in order to describe the participants' search behavior and experiences.

Search Activities

The participants carried out 331 search sessions during the study, comprised of 735 queries. The average length of a search session was 187 seconds with on average 134 seconds being spent in the search interface (Table 1). The participants were slightly more active during the first half of the study with 54% of the search sessions taking place during the first two weeks. There was a high degree of overlap between the queries (primarily within each participant), with 42% of queries being completely unique. The query related figures are similar to those reported in other recent log analysis studies of mobile search (Kamvar, Kellar, Patel, & Xu, 2009). The distribution of query term frequencies followed a long tail pattern, with one and two term queries making up 74% of all queries, which is also congruent with previous studies on mobile Web search patterns (e.g., Church et al., 2008). This suggests that in terms of query characteristics our participants match the broader population of mobile Web users.

As the distribution data in Table 1 shows, there was considerable variance in search activity between participants. Some of this variance is explained by external factors, for example overall lack of mobile information needs during

Table 1. Search activity averages per participant

Search sessions	20 (Md = 13, range = 5 – 72, SD = 17)
Duration of search session	187 seconds (Md = 152, SD = 95)
Time spent in the search interface	134 seconds (Md = 118, SD = 52)
Number of queries	43 (Md = 32, range = 10 – 164, SD = 37)
Queries per session	2.3 (Md = 2, SD = 0.8)
Terms per query	2.1 (Md = 2, SD = 1)

the study period, or the utilization of other mobile information access methods such as Web browsing or application use. Although the average number of search sessions and queries over the four-week period may seem low, they are in line with the figures Kamvar et al. (2009) used to identify "frequent users" (individuals who used Google for at least 10 search tasks over a 35-day period) in their large scale analysis of mobile search.

We asked the participants to estimate the proportion of mobile Web search queries they carried out with the prototype. Just one participant used it in less than one third of queries, three participants in about half of queries, another three in two thirds of their queries and the rest (ten participants in all) in nearly all of their queries. The main reason for switching to other search engines (in practice Google), reported by eight of the frequent clustering interface users, were problems with finding the information they were looking for. Five participants reported having used the clustering interface as the initial search engine and only switching to Google after a few unsuccessful queries. Another reason for choosing Google appears to have been the need for expedited result access – this implies an *a priori* understanding that it would be more effective for the given information need. This suggests that while many participants utilized our interface almost exclusively or treated it as the method of first choice, they were also able to estimate the expected benefits of the clustering interface over other methods, and adjust their information access strategies accordingly.

Category Interactions during Sessions and Queries

The participants made 474 category selections and 189 selections targeting the 'All results' option (AR) during the study. They used categories more actively than AR for results access (Table 2). All of the participants used the categories to access results; however three participants did not utilize the AR option at all during the study.

Categories and AR were used in 95% of sessions, with categories being used in 77% and AR in 37% of all sessions (Table 3). Both

Table 2. Average number of overall, session specific and query specific category list entry selections by result access method

	Categories	**AR**
Average number of selections during the study	28 (Md = 17, SD = 26)	11 (Md = 6, SD = 12)
Average number of selections per session	1.5 (Md = 1, SD = 0.5)	0.5 (Md = 0.4, SD = 0.4)
Average number of selections per query	1.2 (Md= 1.2, SD = 0.2)	0.5 (Md = 0.3, SD = 0.5)

Table 3. Distributions of category interactions during queries and sessions (differences to 100% due to rounding)

	Category selections		**AR selections**	
Number of selections	Percentage of queries	Percentage of sessions	Percentage of queries	Percentage of sessions
0	39%	23%	73%	63%
1	52%	45%	25%	25%
2	6%	17%	< 1%	8%
3	1%	7%	< 1%	2%
4 or more	1%	9%	< 1%	2%

were used in the same session in 19% of cases. At the level of individual queries, categories were used in 61% of the queries where it was an option (i.e., queries that resulted in a category list being shown; 88% of all queries) and the AR in 27% of queries. There was very little overlap in category and AR use at query level (only 6% of queries).

The effect of device type and screen size on the use of categories was investigated but no correlations were found. However, there was a slight overall decreasing trend in the use of categories over time. During the first two weeks of the study the categories were used in 65% of queries and in 55% during the second half of the study. The change is not statistically significant, which is explained by the individual variation in category use: with four participants the use of categories decreased and with two participants increased by 25 percentage points or more from the first to the second half of the study.

We conducted a more detailed analysis of query level interactions. Categories were the first selection subsequent to the query in 58% of the cases, whereas AR was used as the primary option in 22% of queries. A clear behavioral pattern emerged in the transitions between categories an AR during queries where both were used: in 76% of cases the categories were used first, followed by a switch to AR during the same query. In 9% of queries the session continued with a reformulation of the original query without the use of categories or AR. In 11% of cases the query was abandoned (i.e., either session ended or a new query on a different topic followed). There was a statistically significant difference in actions taken following a query, X^2 (3) = 31.33, $p < 0.001$. Post-hoc analysis with Wilcoxon signed-ranks tests was conducted with a Bonferroni correction applied. The participants were more likely to initially use categories than AR ($Z = -3.43$, $p < 0.01$), reformulate the query ($Z = -1.48$, $p < 0.001$) or abandon the query ($Z = -3.62$, $p < 0.001$).

Search Success

In terms of search success, ten participants stated that they were able to find the answers to their information needs always or almost always. Four participants were able to find answers on occasion (in 33% to 66% of the time) and two participants in less than one third of their queries. The participants who were not usually able to find results satisfying their information needs cited problems with the categories as one of the key reasons. Those participants that did usually find satisfactory results using the clustering interface also had some concerns; for example, having to resort to AR on occasion when the categories failed to contain suitable results, or having to access multiple categories before finding suitable results.

One method of quantifying search success is the query click-thru rate. It examines in how many queries the users followed at least one search result. The participants had an average

click-thru rate of 47% after using the categories and 51% from the ranked result list. The difference in click-thru rates is not significant. However, an interesting phenomenon was observed when investigating situations where more than one result is selected per query. When categories are used, in 11% of the queries leading to a result selection our participants selected two or more results. When AR was used, the corresponding figure is 7%. It is noteworthy that nine participants made multiple selections per query when using the categories, and only four participants did so when using the ranked result list. Although it is not feasible to carry out significance testing because of the low number of participants involved, this does suggest that categories appear to support result exploration better than the ranked result list.

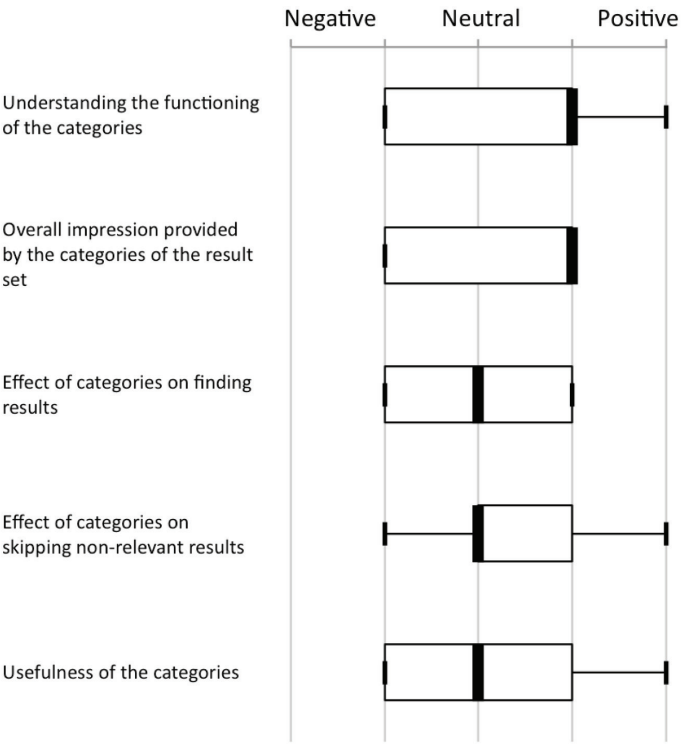
Subjective Assessment of the Search Result Categories

Subjective feedback on the search result categories was collected with a questionnaire and a related semi-structured interview. In the questionnaire, the participants were presented with five claims about category use with a 5-point, bipolar response scale (Figure 2). The participants considered it to be relatively easy to understand the functioning of the clustering which is unsurprising given the focus on the most common words and phrases. The categories also provided a fairly informative overall impression of the whole search result set. The participants' assessment was neutral in regard to the effect of the categories on finding useful results, bypassing non-relevant results and general usefulness. We further examined the relationship between perceived usefulness and the potentially beneficial features of clustering, the provision of an overview and ability to find actionable results. We found that the perceived usefulness of categories strongly correlated with the overall impression provided of the result set (Spearman's rank correlation coefficient, $r_s = 0.89$, $N = 17$, $p < 0.001$) and effect of categories on finding results ($r_s = .67$, $N = 17$, $p < 0.01$). Interestingly, mobile device features such as screen size or interaction type (touchscreen versus keypad) did not correlate with the subjective assessments, indicating that the search interface was the main contributor.

Interview responses provided more insights into the notion of usefulness. According to the participants, clustering interface provides *situational* benefits, specifically on queries where the information need is difficult to identify, either due to search inexperience or limited understanding of the topic area. Five participants noted that the categories help if one struggles with specifying the information need; categories aid in *narrowing down the search into a specific set of results* and making better searches by *suggesting better keywords* to use in the queries. There are several instances of the former strategy in the interaction data, especially for informational queries. For example, the query "gremlins" resulted in a category selection "gremlins movie," followed by a result selection that took the participant to the topic listing for the movie at Answers.com. The query data also supports the latter observation; four participants reformulated their queries using terms from the category list. A more prevalent reformulation strategy, exhibited by nine participants at least once, was to execute a new query on the same topic after being presented with the categories, sometimes several times, before selecting a category (or AR) and then accessing a search result. This behavior suggests that the categories were used diagnostically to evaluate the success of the query before committing to result access.

When inquired about the suitability of the clustering categories for their search strategies, seven participants stated that the categories helped in finding relevant results. However, the outcomes were overall mixed, with the usefulness of the categories varying between queries. Two of these participants mentioned that short queries seemed to result in better category labels, and one of these participants started intentionally shortening queries by one term and then picking the most promising category from the list. One example of this behavior entails the participant issuing the query "charcoal maker,"

Figure 2. Subjective rating of search result categories (boxplots show median, 1st and 3rd quartiles and range of values)



and selecting the category leading to results related to the pasta dish with the same name.

The interview comments by our participants further suggest that more experienced searchers may not find the category-based assistance as useful, and *categories make the process more laborious if the user knows what she is looking for*. In these cases the user expects to find the result among the first few results, and the category interface necessitates an additional click before the user gets to the result list. In addition, the quality of cluster labels was mentioned as a detractor. Six participants discussed the lack of descriptiveness of the category labels, which made it difficult to form an understanding of the result set. Two of these participants noted that some form of preview of the category content would have been useful. Such previews would not only aid in assessing the relevance of the category when the label is not descriptive enough, but also serve as direct links to likely results of interest without having to browse into the category.

We further inquired about how our participants understood the clustering process, and which structures other than most frequent terms and phrases the participant would find useful. Four participants explained that the algorithm *should understand their intent*. For example, when searching with the query "student allowance," the categories should have contained entries related to the payment schedule, application process and so on; similarly to how this information would be broken down in a human-generated tuition guide. Nine participants commented on the importance of *location-based results*. Another common theme, mentioned by seven participants, was

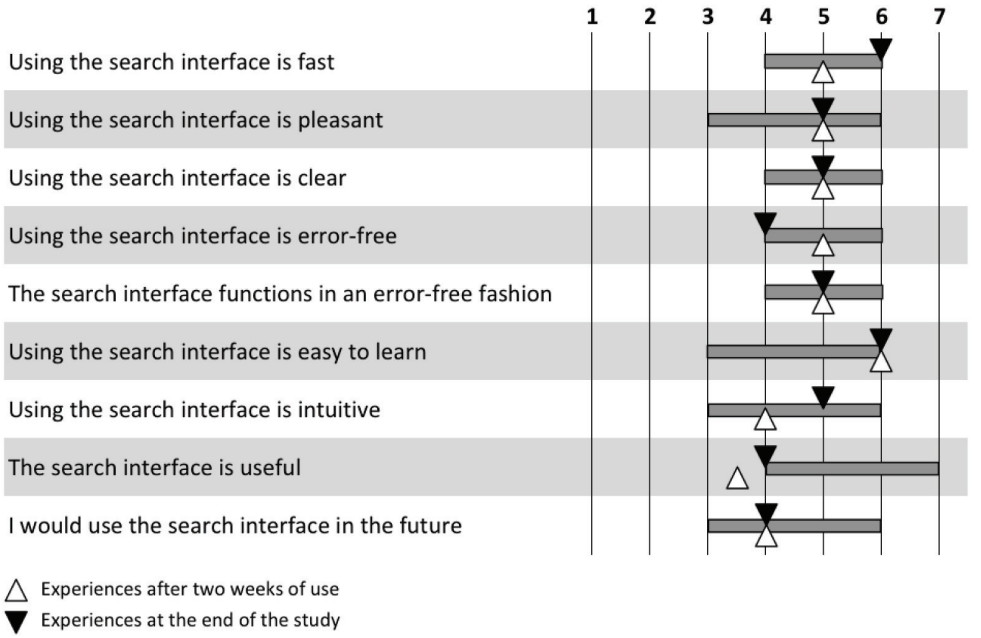
organization of results by *content type*, which would allow one to either drill into or filter out certain type of results.

Expectations and Perceived User Experience

The participants reported their initial experiences with the search prototype after two weeks of use (Figure 3). The results indicate that the participants' initial experiences fell within the range of their expectations in most cases; only for perceived usefulness their experiences were slightly negative (median = 3.5). For ease of learning the situation was the opposite, with experiences meeting the highest desired level (median = 6). At the end of the study most of the experience dimensions were rated within the range of expectations at a slightly (5) or moderately (6) positive level. Only error-free use, usefulness and future are rated as neutral (4), the lowest acceptable level. In all other cases except for perception of error-free use, the change from initial to the final experiences was positive. The difference between the intermediate and final assessment is significant for error-free use (within-subject Wilcoxon signed-ranks test; $Z = -2.22$, $p < 0.05$) and usefulness ($Z = -2.44$, $p < 0.05$). As with other subjective assessments, mobile device features did not correlate with the user experience ratings.

After completing the post-study questionnaire, participants listed the aspects they considered to be essential for good user experience in a mobile Web search interface. Both the high and low rated experience dimensions are reflected in their preferences. The most frequently mentioned important features include "speed of use" (16 participants), "usefulness" (13), "clarity" (13), and "error-free function" (10). The word selections agree with these rankings. Positive terms included terms such as "easy to use," "fast," "reduced," "simple," and "useful," and the top negative words included terms such as "clumsy," "time consum-

Figure 3. SUXES measures: range of expectations (gray line), experiences after 2 weeks (white marker), and post-study experiences (black marker); all values represent median



ing," "confusing," and "conflicting." The overall word associations were positive, with 59% of the top ranking word selections consisting of positive terms. The difference between the positive and negative word selections is not statistically significant.

These quantitative measures of user experience complement the open-ended comments by the participants. They would appreciate an uncluttered and streamlined search experience that provides tangible value, with an interface that is easy to learn and use without errors. The clustering interface was able to provide a positive search experience in many respects; however its use and usefulness suffered from shortcomings in the proposed clustering approach. The significant increase in perceived usefulness over time suggests that the participants might have learned to better utilize the categories with an even prolonged exposure to the system.

DISCUSSION

In this study we sought to answer two research questions that focus on the use and utility of clustering categories as a presentation and interaction method in the mobile Web search interface. Towards this end, we first discuss the interaction patterns we identified in realistic mobile information access scenarios. Next, we examine in which situations category-based access can benefit users and what the limitations are. Finally, based on the user feedback gained during the study and observations made of the interaction data, we propose design considerations for category-based mobile search interfaces.

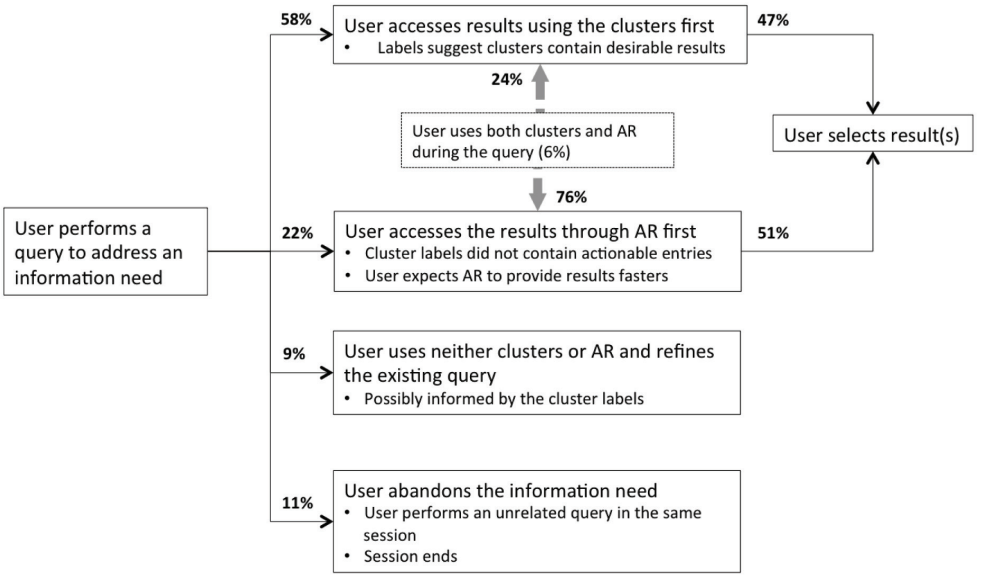
Use of Categories for Search Result Access

The analysis of the longitudinal interaction data shows that categories were actively used for result access, whereby the entries in the category list were utilized in 77% of sessions and 61% of queries. Our data shows that when the category list was available, in most cases our participants used it as the primary means for accessing the results over the ranked result list. These quantitative observations were reflected in the interview responses, with 53% of participants stating they used the categories always or almost always when searching with the prototype. When the category list failed to provide actionable items to pursue, the participants either used the ranked result list to access the results or reformulated their query. However, categories were not able to support searches in all situations, as in about 11% of queries they abandoned the search completely at the category list stage without initiating other interactions. Based on these observations, we identified the clustering interface search model depicted in Figure 4.

The use of categories and ranked result list led to relatively high click-thru rates. As the figures are based on queries where categories were utilized, it is not possible to make direct comparisons to previous studies. Nevertheless, our data appear to be closer to desktop search click-thru rates (Käki, 2005a) than those reported of large-scale mobile search patterns (e.g., Church et al., 2008). Although the present study focused on the active mobile search user population, we find it unlikely that it would be the only contributing factor for this difference. One potential explanation is that the category-based interface helped the participants choose the results access method most suitable for their information need (i.e., use categories, use AR or reformulate query). This in turn may have reduced the number of subsequent failed query reformulations, and search results missed due to lack of exploration and information overlook.

In addition to overall interaction patterns, we saw some evidence of categories supporting explorative search. Category use resulted in more results being selected when more than one result was followed. Our interpretation is that the participants intentionally chose to use the categories when they had information needs that required researching multiple sources to identify the answer. This is similar to the findings on desktop search by Käki (2005a), who found that people were more likely to access

Figure 4. Participants' search behavior with the clustering interface



multiple results for a query through categories than the ranked result list. In addition, the participants usually restricted their access to a single category (85% of cases), which suggests that the result clusters were generally able to provide relevant subtopics to drill into without the need to access multiple categories.

Although the categories were the more popular method of browsing the results than the ranked result list, there are some factors that may have influenced this behavior. Participant feedback suggests that there was sometimes an element of trial and error to the use of the categories due to non-descriptive labels, leading to multiple category selections. Additionally, the category-driven presentation approach may also have influenced the participants' choice of result access. For example, in the desktop study by Käki (2005a), categories were utilized only in about 26% of queries. However, in the desktop interface the categories were presented next to the ranked result list. In our interface only the category list was shown initially, which might have led the participants to use the clusters more than they otherwise would have in some situations.

Benefits and Limitations of Category-based Result Access

Our results suggest similar benefits of category-based result access that have been identified in previous studies, but also show that in real information seeking scenarios it is vital to also consider the effect of the mobile context, and the users' expectations towards the category construction process with respect to their existing search strategies.

The feedback from participants highlights several useful features of clustering categories. For example, they can be useful in narrowing down the search into a subtopic of the results, supporting the findings of previous experiments (Carpineto et al., 2009a; Heimonen, 2008). Cluster labels can also aid by suggesting new, content-based query terms, as proposed by Carpineto et al. (2009a). This can be helpful in dealing with hard to describe information needs, which are typical situations where the search engine can fail to identify the user's information need and provide good results on the first page. These failure situations have been identified as one success case for alternative presentation

methods (Carpineto et al., 2009b; Jones et al., 2004; Käki, 2005a). Addressing challenging and vague information needs is especially critical for mobile search given the pronounced lack of result exploration. Clustering can alleviate the situation by exposing the user to potentially useful results beyond the first result page.

However, the usefulness of clustering categories is influenced by several factors; including the information needs themselves and the quality of clustering outcomes. The type and distribution of information needs is likely the most important factor when considering real world applicability of the category-based approach. The analysis by Church et al. (2008) shows that approximately 30% of mobile searches are *navigational*, while the rest are *informational* (10%) or *transactional* (60%). While clustering is likely to be most useful for informational queries, it could also assist in filtering of and choosing between different content types (e.g., videos or images) or metadata (e.g., reviews versus specifications of a product) for transactional queries. Käki (2005a) found that categories were less likely to be used for addressing navigational queries on the desktop, which also likely holds for mobile search. This in part explains the situational usefulness of the categories mentioned by our participants.

Our findings also indicate that categories were not particularly suitable to answering focused, well-defined information needs. In such cases our participants anticipated to find the intended result within the top ranking results returned by the search engine. Having to make a selection to access the ranked result list was an unnecessary extra step. Clearly, the interface design of category-based search interfaces should aim at addressing both focused and more ambiguous information needs. As Carpineto et al. (2009b) emphasize, category-based access should be viewed as a complementary rather than a completely alternative approach to the traditional interfaces. However, more research is clearly warranted in order to study the connections between mobile information needs, category use and users' preferred search strategies.

Cluster label descriptiveness, consistency and predictability have a fundamental effect on how the category interface is perceived and used. Although our clustering algorithm utilized the most common words and phrases in result captions, it did not consistently provide understandable labels. This resulted in situations where the participants had problems identifying categories that would contain satisfactory results, which in turn negatively impacted the user experience. Similar findings related to label quality have been reported in previous studies (Kules & Shneiderman, 2008). Improving both the representativeness of the labels and their conceptual coverage of the search topic is necessary in order to increase the usefulness of the categories and minimizing unnecessary navigation caused by uninformative labels.

Design Guidelines for Category-based Mobile Search Interfaces

The results of the study contributed to two design guidelines related to the design of category-based mobile search interfaces. First, one should consider how and when the category overviews should be included in the search interface. Second, one also needs to consider the options for producing the categories to address different kinds of information needs and user expectations.

Consider the Presentation and Content of Category Overviews

The category interface should not impose an added cognitive burden on the user and it should work to its main strength, which is providing a clear and useful overview of the search results. This can be broken down into specific design challenges: *how* the category overview is presented and interacted with, *when* should the categories be shown, and *which* information to include in the category overview. Previous interface designs have proposed "category-driven" approaches to address the first challenge, whereby the list of category labels is provided as the initial results view (e.g., Carpineto et al., 2009a; De Luca & Nürnberger,

2005; Heimonen & Käki, 2007; Machado et al., 2009). The situational benefits of categories and the navigation issues caused by inconsistent labeling necessitate a rethink of the category-driven approach. We propose a solution where the category overview includes the top results from the ranking result list (Figure 5).

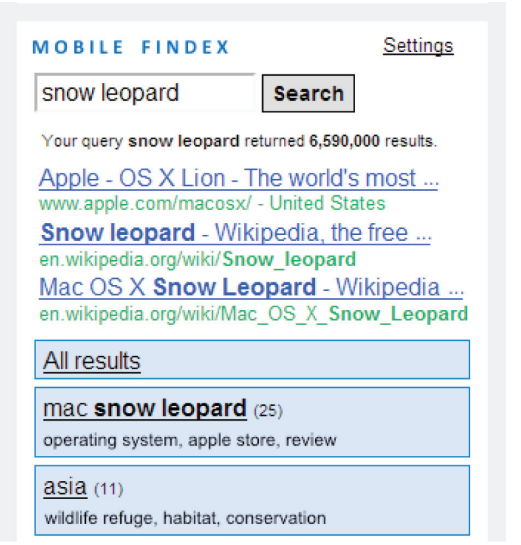
This solution has the benefit of supporting different information needs and search strategies simultaneously. Users who are satisfied with the top ranking results do not need to utilize the categories, and in cases where subtopic exploration of the results is required, the category list is available. However, one must also consider the user's preferences with respect to how the search interface is organized. While categories can be useful, some users might prefer the traditional ranked result list as the primary interface. As suggested by some of our participants, the system could adopt a "results-driven" design, whereby access to the category overview can be requested by the user on-demand. Some form of explicit control by the user is likely necessary for triggering the category view, given the poor reception of systems that try to automatically infer the user's intent (Hearst, 2009).

In addition to providing a usable presentation and interaction method for accessing the categories, their content should also be as informative and appropriate as possible. While attempting to improve the labeling algorithms themselves is a necessary endeavor, it may also beneficial to supplement the labels with additional descriptive elements. Content-based keyphrases (Jones et al., 2004) could be a useful addition to the cluster labels in describing the content of the result category, especially in situations where the cluster label is ambiguous (see Figure 5 for an example). However, the quality of the source material limits the applicability of any purely content-based approaches and other sources should also be considered. When available, user-generated metadata could be useful. For example, Carmel, Roitman, and Zwerdling (2009) had success with improving cluster labels using terms extracted from Wikipedia.

Consider Different Categorization Approaches and Contextual Factors

In this study we utilized a clustering algorithm that created the clusters based on the most common words and phrases in the result captions.

Figure 5. Revised category list view with top ranking results and enhanced category labels



Although this was found useful in some cases, the need for more sophisticated approaches was also identified. This is necessary not only to match the users' expectations for the categorization process but also to account for contextual factors.

First, there is the issue of clustering outcomes being less predictable and consistent than those of classification-based category systems. The clustering approach employed in the prototype disagreed with the conceptual search strategies of some of our participants, who stated they would have preferred an algorithm that "understands" their search intent better. Thus, instead of operating at the level of words and phrases, the algorithm should have extracted topical terms and concepts relevant to the user's intended query target. This likely requires some form of topic specific, human-generated category structure to be applied based on the query intent. However, providing a single category structure may not always be applicable. Supporting multiple types of categories, or even the creation of one's own categories, could allow the users to utilize the representation that best fits their needs (Kules & Shneiderman, 2008). The challenge here is to match the user's information need and expectations with the correct representations, which can be especially challenging in the Web search domain.

Second, the contextual nature of mobile information needs should be accounted for in the category construction process. Contextual aspects, such as location and activity, can trigger unique information needs and affect how they are fulfilled. Unsurprisingly, many of our participants mentioned having location-based information needs. Previous research suggests that location-based mobile search could benefit from features that enable iterative, exploratory and comparative search activities (Amin et al., 2009). Plain content-based clustering is unlikely to be efficient in capturing and highlighting these contextual aspects. One way to make clustering approaches context-aware would be to utilize the contextual metadata of results, such as the geographical location or inferred social activity (e.g., an ongoing local event) as measures of similarity. Alternatively, these metadata could be used to introduce a filtering scheme on top of the clusters (e.g., only show results related to the user's current location). Provided the query intent can be accurately identified and matched with an existing representation, faceted browsing (Karlson et al., 2006) could also support lookup of local service and business information using consistent and familiar information structures.

Limitations and Future Research

Generalizing the results of this study has certain limitations because of the limited sample size and focus on active mobile search users. Studying less active and inexperienced users would be an interesting future research area to ascertain how well clustering supports their needs and search strategies.

In user studies one can expect to encounter methodological concerns such as the impact of demand characteristics and the expectation that the results should establish the worth of the system being studied (Brown, Reeves, & Sherwood, 2011). We believe that in this study the main effect of demand characteristics was an increased likelihood of experimenting with the clustering categories at the start of the study. However, it is also likely that the length of the user study and the perceived usefulness of clustering across information needs balanced this effect out. In addition, the participants were very candid in their feedback, which provided a very well rounded perspective of the benefits and drawbacks of both clustering and the search interface. The second concern highlighted by Brown et al. (2011) is the tendency for user trials to aim at proving the superiority of the evaluated design over its alternatives. Our aim was instead to focus on the lessons we can learn from the study in order to better understand the benefits and drawbacks of clustering as a search interface paradigm.

In addition to existing search approaches, altogether new ways of mobile information access are emerging. For example, Jones

(2011) argues that in the future mobile search interfaces should consider a wider set of search scenarios beyond the traditional fast-paced keyword-based lookup. One recent example of such approaches is the emergence of social search interfaces that emphasize other users' queries and questions in the search process (Arter et al., 2007; Church et al., 2010). We see opportunities for category-based interfaces to support these ways of searching by providing means of organization, filtering and browsing of the social content, for example by highlighting common query topics.

CONCLUSION

We conducted a longitudinal study with 17 participants that examined the use of clustering categories for result access in a mobile Web search interface prototype. The aim of the study was to find out how the categories are utilized and in which situations they are found to be useful. Although the clustering categories were used as the primary method of result access, the subjective feedback on their usefulness was mixed. Our results show that categories can be situationally helpful, for example when one is unsure of how to frame the information need and requires guidance on how to narrow the search, or when the information need itself cannot be expressed with a focused query that would provide the desired item within the first search results. The usefulness of category-based result access is influenced not only by the nature of the information needs, but also by the category generation method and how the categories are presented and interacted with in the search interface.

The results suggest several design implications for category-based search interfaces. An interface that combines top ranking results and categories would facilitate both navigation and lookup oriented query intents and explorative search. Additionally, the presentation of the categories should aim at providing more consistent and descriptive overviews, for example by including cluster content previews. Further, we identified the need for more intelligent category creation approaches that would account for the contextual aspects of mobile information needs and use familiar representations matching the user's expectations. Approaches that include both clustering and human-generated classifications while also incorporating contextual cues are likely to be beneficial across a wider variety of mobile information needs than clustering approaches utilizing only on the textual content of the result captions.

ACKNOWLEDGMENT

Appreciation is extended to Kari-Jouko Räihä and the anonymous reviewers for their thoughtful comments, and to Mika Käki for the use of the Findex search and clustering framework for the purposes of this study. This research was financially supported by the UCIT Graduate School in User-Centered Information Technology.

REFERENCES

Amin, A., Townsend, S., Ossenbruggen, J., & Hardman, L. (2009). Fancy a drink in Canary Wharf? A user study on location-based mobile search. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I* (pp. 736-749). Berlin, Germany: Springer-Verlag.

Arter, D., Buchanan, G., Jones, M., & Harper, R. (2007). Incidental information and mobile search. In *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 413-420). New York, NY: ACM.

Benedek, J., & Miner, T. (2002, July). *Measuring desirability: New methods for evaluating desirability in a usability lab setting, a paper on the Desirability Toolkit.* Paper presented at the Usability Professionals Association Conference, Orlando, FL.

Brown, B., Reeves, S., & Sherwood, S. (2011). Into the wild: Challenges and opportunities for field trial methods. In *Proceedings of the Annual Conference on Human Factors in Computing Systems* (pp. 1657-1666). New York, NY: ACM.

Buchanan, G., Jones, M., & Marsden, G. (2002). Exploring small screen digital library access with the Greenstone Digital Library. In *Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries* (pp. 583-596). Berlin, Germany: Springer-Verlag.

Carmel, D., Roitman, H., & Zwerdling, N. (2009). Enhancing cluster labeling using Wikipedia. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 139-146). New York, NY: ACM.

Carpineto, C., Della Pietra, A., Mizzarro, S., & Romano, G. (2006). Mobile clustering engine. In M. Lalmas, A. MacFarlane, S. Rüger, A. Tombros, T. Tsikrika, & A. Yavlinsky (Eds.) *Proceedings of the 28th European Conference on Information Retrieval Research* (LNCS 3936, pp. 155-166). Berlin, Germany: Springer-Verlag.

Carpineto, C., Mizzaro, S., Romano, G., & Snidero, M. (2009a). Mobile information retrieval with search results clustering: Prototypes and evaluations. *Journal of the American Society for Information Science and Technology*, *60*(5), 877–895. doi:10.1002/asi.21036

Carpineto, C., Osiński, S., Romano, G., & Weiss, D. (2009b). A survey of Web clustering engines. *ACM Computing Surveys*, *41*(3), 17. doi:10.1145/1541880.1541884

Chen, H., & Dumais, S. (2000). Bringing order to the web: Automatically categorizing search results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 145-152). New York, NY: ACM.

Church, K., Neumann, J., Cherubini, M., & Oliver, N. (2010). SocialSearchBrowser: A novel mobile search and information discovery tool. In *Proceedings of the 15th International Conference on Intelligent User Interfaces* (pp. 101-110). New York, NY: ACM.

Church, K., & Oliver, N. (2011). Understanding mobile web and mobile search use in today's dynamic mobile landscape. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 67-76). New York, NY: ACM.

Church, K., & Smyth, B. (2009). Understanding the intent behind mobile information needs. In *Proceedings of the 13th International Conference on Intelligent User Interfaces* (pp. 247-256). New York, NY: ACM.

Church, K., Smyth, B., Bradley, K., & Cotter, P. (2008). A large scale study of European mobile search behaviour. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 13-22). New York, NY: ACM.

Church, K., Smyth, B., & Keane, M. T. (2006). Evaluating interfaces for intelligent mobile search. In *Proceedings of the International Cross-Disciplinary Workshop on Web Accessibility* (pp. 69-78). New York, NY: ACM.

Cutting, D., Karger, D., Pedersen, J., & Tukey, J. (1992). Scatter/Gather: Aa cluster-based approach to browsing large document collections. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 318-329). New York, NY: ACM.

De Luca, E. W., & Nürnberger, A. (2005). Supporting information retrieval on mobile devices. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 347-348). New York, NY: ACM.

Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM*, *49*(4), 59–61. doi:10.1145/1121949.1121983

Hearst, M. A. (2009). *Search user interfaces*. New York, NY: Cambridge University Press.

Heimonen, T. (2008). Mobile Findex: Facilitating information access in mobile web search with automatic result clustering. *Advances in Human-Computer Interaction*, *2008*, 680640. doi:10.1155/2008/680640

Heimonen, T. (2009). Information needs and practices of active mobile Internet users. In *Proceedings of the 6th International Conference on Mobile Technology, Applications & Systems* (p. 50). New York, NY: ACM.

Heimonen, T., & Käki, M. (2007). Mobile Findex - supporting mobile web search with automatic result categories. In *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 397-404). New York, NY: ACM.

Hinze, A., Chang, C., & Nichols, D. M. (2010). Contextual queries express mobile information needs. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 327-336). New York, NY: ACM.

Jones, M. (2011). Classic and alternative mobile search: a review and agenda. *International Journal of Mobile Human Computer Interaction*, *3*(1), 22–36. doi:10.4018/jmhci.2011010102

Jones, M., & Marsden, G. (2006). *Mobile interaction design*. New York, NY: John Wiley & Sons.

Jones, S., Jones, M., & Deo, S. (2004). Using keyphrases as search result surrogates on small screen devices. *Personal and Ubiquitous Computing*, *8*(1), 55–68. doi:10.1007/s00779-004-0258-y

Kaikkonen, A. (2011). Mobile internet, internet on mobiles or just internet you access with variety of devices? In *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (pp. 173-176). New York, NY: ACM.

Käki, M. (2005a). Findex: Search result categories help users when document ranking fails. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 131-140). New York, NY: ACM.

Käki, M. (2005b). *Enhancing web search result access with automatic categorization* (Unpublished doctoral dissertation). Retrieved from http://acta.uta.fi/english/teos.php?id=10797

Käki, M. (2006). fKWIC: Frequency-based keyword-in-context index for filtering web search results. *Journal of the American Society for Information Science and Technology*, *57*(12), 1606–1615. doi:10.1002/asi.20338

Kamvar, M., & 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* (pp. 701-709). New York, NY: ACM.

Kamvar, M., Kellar, M., Patel, R., & Xu, Y. (2009). Computers and iPhones and mobile phones, oh my! A logs-based comparison of search users on different devices. In *Proceedings of the 18th International Conference on World Wide Web* (pp. 801-810). New York, NY: ACM.

Karlson, A., Robertson, G., Robbins, D., Czerwinski, M., & Smith, G. (2006). FaThumb: a facet-based interface for mobile search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 711-720). New York, NY: ACM.

Kules, B., & Shneiderman, B. (2008). Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management*, *44*(2), 463–484. doi:10.1016/j.ipm.2007.07.014

Machado, D., Barbosa, T., Pais, S., Martins, B., & Dias, G. (2009). Universal mobile information retrieval. In C. Stephanidis (Ed.), *Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environments* (LNCS 5615, pp. 345-354). Berlin, Germany: Springer-Verlag.

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (pp. 1-7). New York, NY: ACM.

Sohn, T., Li, K. A., Griswold, W. G., & Hollan, J. D. (2008). A diary study of mobile information needs. In *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (pp. 433-442). New York, NY: ACM.

Taylor, C. A., Anicello, O., Somohano, S., Samuels, N., Whitaker, L., & Ramey, J. A. (2008). A framework for understanding mobile internet motivations and behaviors. In *Proceedings of the Extended Abstracts of the International Conference on Human Factors in Computing Systems* (pp. 2679-2684). New York, NY: ACM.

Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., & Hella, J. (2009). SUXES - user experience evaluation method for spoken and multimodal interaction. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association* (pp. 2567-2570). Baixas, France: ISCA.

White, R. W., Drucker, S., Kules, B., & Schraefel, M. C. (2006). Supporting exploratory search. *Communications of the ACM*, *49*(4), 36–39. doi:10.1145/1121949.1121978

White, R. W., Muresan, G., & Marchionini, G. (2006). Report on ACM SIGIR 2006 workshop on evaluating exploratory search systems. *SIGIR Forum, 40*(2), 52-60.

Wilson, M., & Schraefel, M. C. (2008). A longitudinal study of exploratory and keyword search. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 52-56). New York, NY: ACM.

Tomi Heimonen is a researcher and PhD student at the School of Information Sciences at the University of Tampere. He has worked with design, implementation and evaluation issues of interactive systems since 1999 at the Tampere Unit for Computer-Human Interaction, and has published in the areas of information visualization, mobile information access and multimodal interaction. His current research interests include multimodal mobile interfaces, gesture-based interaction and pervasive interactive spaces. His PhD research is focused on designing and studying novel search interface solutions for mobile search.

CALL FOR ARTICLES

International Journal of Mobile Human Computer Interaction

An official publication of the Information Resources Management Association

The Editor-in-Chief of the *International Journal of Mobile Human Computer Interaction* (IJMHCI) would like to invite you to consider submitting a manuscript for inclusion in this scholarly journal.

ISSN 1942-390x
eISSN 1942-3918
Published quarterly

MISSION

The primary objective of the *International Journal of Mobile Human Computer Interaction* (JMHCI) is to provide comprehensive coverage and understanding of the issues associated with the design, evaluation, and use of mobile technologies. This journal focuses on human-computer interaction related to the innovation and research in the design, evaluation, and use of innovative handheld, mobile, and wearable technologies in order to broaden the overall body of knowledge regarding such issues. IJMHCI also considers issues associated with the social and/or organizational impacts of such technologies.

COVERAGE/MAJOR TOPICS

- Case studies and/or reflections on experience (e.g. descriptions of successful mobile user interfaces, evaluation set-ups, etc.)
- Context-aware/context-sensitive mobile application design, evaluation, and use
- Design methods/approaches for mobile user interfaces
- Ethical implications of mobile evaluations
- Field-based evaluations and evaluation techniques
- Gestural interaction techniques for mobile technologies
- Graphical interaction techniques for mobile technologies
- Issues of heterogeneity of mobile device interfaces/interaction
- Lab v. field evaluations and evaluation techniques
- Lab-based evaluations and evaluation techniques
- Mobile advanced training application design, evaluation, and use
- Mobile assistive technologies design, evaluation, and use
- Mobile commerce application design, evaluation, and use
- Mobile HCI lab design/set-up
- Mobile healthcare application design, evaluation, and use
- Mobile interactive play design, evaluation, and use
- Mobile learning application design, evaluation, and use
- Mobile technology design, evaluation, and use by special (needs) groups (e.g. elderly, children, and disabled)
- Multimodal interaction on mobile technologies
- Non-speech audio-based interaction techniques for mobile technologies
- Other emerging interaction techniques for mobile technologies
- Other related issues that impact the design, evaluation, and use of mobile technologies
- Speech-based interaction techniques for mobile technologies
- Tactile interaction techniques for mobile technologies
- Technology acceptance as it relates to mobile technologies
- User aspects of mobile privacy, security, and trust
- User interface architectures for mobile technologies
- User interface migration from desktop to mobile technologies
- Wearable technology/application and interaction design, evaluation, and use

All submissions should be emailed to:
Joanna Lumsden
Editor-in-Chief
ijmhci@igi-global.com

Ideas for Special Theme Issues may be submitted to the Editor-in-Chief.

Please recommend this publication to your librarian. For a convenient easy-to-use library recommendation form, please visit: http://www.igi-global.com/ijmhci