

Contextual Queries express Mobile Information Needs

Annika M. Hinze
University of Waikato
Hamilton, New Zealand
hinze@cs.waikato.ac.nz

Carole Chang
University of Waikato
Hamilton, New Zealand
sc200@waikato.ac.nz

David M. Nichols
University of Waikato
Hamilton, New Zealand
dmn@cs.waikato.ac.nz

ABSTRACT

The users of mobile devices increasingly use networked services to address their information needs. Questions asked by mobile users are strongly influenced by contextual factors such as location, conversation and activity. We report on a diary study performed to better understand mobile information needs. We find that the type of questions recorded by participants varies across their locations, with differences between home, shopping and in-car contexts. These variations occur both in the query terms and in the form of desired answers. Both the location of queries and the participants' activities affected participants' questions. When information needs were affected by both location and activity, they tended to be strongly affected by both factors. The overall picture that emerges is one of multiple contextual influences interacting to shape mobile information needs. Mobile devices that attempt to adapt to users' context will need to account for a rich variety of situational factors.

Categories and Subject Descriptors

H.5.m [Information Interfaces and Presentation]: e.g., HCI

General Terms

Design

Keywords

Diary Study, user requirements, mobile information needs, context, location

1. INTRODUCTION

The popularity of mobile devices makes it now common to see people address their information needs in-situ, using services such as web search or digital maps. The growth of mobile activities has been accompanied by devices that can sense more features of their environment, including location, orientation, temperature and lighting. Designers combine this ambient information with other data to enable devices to adapt their capabilities to the user's context. In this paper, we examine the information needs of mobile users to inform the design of mobile information search applications. Information needs are, by their nature, ubiquitous: they can arise on any topic, at any time and in any location. Previous work shows that many of these mobile information needs are never satisfied, often being postponed and abandoned [17]. Context-adaptation

provides a potential mechanism for portable devices to bridge this gap between actual and satisfied needs. We use a diary study to investigate how contextual factors affect mobile information needs and how they can be exploited to enhance application effectiveness.

Section 2 outlines previous work in the area. We then describe our study methodology in Section 3, followed by an analysis of the results from several perspectives in Section 4. In Section 5, we discuss the implications of the results for context-aware applications. We conclude with the key contributions and ideas for future work.

2. BACKGROUND

Variable elements of contexts include location, time, social situations and the connectivity or services available from the device [9]. New device functionality enables users to change their behaviour, to plan activities less (as connectivity enables dynamic re-planning), to rely on device-maps (rather than paper ones) and to use their time in different manners (such as device use when commuting). When mobile users become accustomed to ubiquitous connectivity, they have the potential to satisfy their information needs through mobile web search (e.g. [1, 10]).

To understand mobile users' information search behaviour, researchers have followed two main paths: analysis of logs of actual mobile search queries (e.g. [10, 4]), and user-centred approaches such as diaries and interviews [17, 4, 1, 6, 9, 14]. Query logs are a valuable resource but the "numbers don't tell the story behind a user's experience—we know for what and when a user queried, but have no context for what inspired the search" [10]. To complement log analysis results, small-scale qualitative methods have been used to study the context of users' mobile queries (e.g. [17, 4, 1, 6, 13, 14, 7]).

The importance of understanding the details surrounding mobile device use is highlighted by Sohn et al.'s [17] finding that 72% of their participants' information needs "were prompted by some contextual factor". Most of the participants in their study also admitted that their desire to satisfy their information need had led them to multi-task their device use with driving a vehicle. Church & Smyth noted that geographical terms increased significantly when their participants were 'mobile' [4].

Lee et al. [12] reported that "most of the time mobile Internet services were used in a public place, without any social interaction, while not moving, and when off duty", whereas more recent work emphasises the social influences on users' usage (e.g. [17, 18, 6, 13]). Amin et al. [1] report that 2/3 of their participants' (location-based) queries were to satisfy a "spontaneous need" and 75% were made in social situations.

Several studies [1, 4, 6, 17, 14, 7] report that location context had a significant impact on their participant's mobile information needs. Location can influence use in several ways: the choice of

which device to use [14], the geographical terms used in queries [4] and the geographical areas most likely to be query targets [1]. Location context can be the most important factor driving contextual information needs [17], however Nylander et al. [14] report that the classic example of “finding out something about the location you are in” represents only 15% of their data.

Mobile query studies often group queries into broad categories, e.g. ‘information seeking’, ‘communication’ and ‘content’ [5]; ‘information seeking’ is then further divided into ‘fact-finding’, ‘information gathering’ and ‘casual browsing’. These categories vary across studies, for example: ‘informational’, ‘geographical’ and ‘personal information management’ [4] and ‘fact-finding’, ‘information gathering’ and ‘non-goal oriented’ [1]. However other studies use either more fine-grained or orthogonal categorisations [14, 17].

The distribution of mobile search topics reported in other studies is difficult to compare as different categories have been used; frequent topics typically include: news, travel & transport, entertainment, trivia, shopping and food & drink [11, 14, 17, 4, 1, 6]. Mobile query log analyses also record ‘adult’ searches [11, 10], which appear not to be reported in the qualitative studies; illustrating the value of using a variety of research methods.

Most of the studies agree that context and location matters, though they do not necessarily agree on their particular influence. A significant number of queries are found to be context-free (i.e., fact-finding) queries. The query topics that were identified in the studies vary, but a high degree of commonality can be observed. Overall, there is a lot of variation in the results that have been obtained by the studies. It is often hard to tell whether the differences between the studies are due to differing participant groups, or study methodology used for the study organisation, or the analytic approach [7].

3. STUDY METHODOLOGY

We performed a small pilot study and a full user study, both using paper diaries. The main difference of our study to the ones reported above is that we did not use the support of electronic devices, such as mobile phones or laptops.

Pilot study.

We used a small pilot study to explore our initial ideas. We devised a diary study to record date and location of the question, the query itself, the type of question, and whether or not the query is related to the current location. For the query type, we distinguished between (a) questions for which the participant needs the answer immediately and the answer affects the participant’s activity (for example, when driving to a restaurant and asking for the nearest free car park), and (b) questions that express the participant’s curiosity about something they saw, where the answer does not necessarily affect their current activity (for example, when sharing a cake with friends to enquire about the recipe of the cake).

The results of the initial study with three users encouraged us to do a diary study with more participants. We used the results of the pilot to improve the diary design: We observed that the relationship to the place cannot be adequately answered in a Boolean fashion. We instead introduced a 5-point scale indicating the strength of relationship between the question and the place. The question about query type seemed to be unclear and did not find clear agreement about the semantics. In the full study we therefore translated it into a question about the relationship between query and activity, thus moving the determination of the type of question from participant to analyst. We also reduced the size of the diary pages to pocket size (8.5 *14.5 cm) for ease of use.

The image shows a completed diary page with the following content:

- Date:** 01/07.
- What is your question?:** How old is the oldest known Maori artefact?.
- Where did you have this question?:** Waitato Museum.
- How strongly is the question related to the place where you had the question?:** A 5-point scale from 'Not at all' to 'The place caused the question'. The scale has five radio buttons, with the fifth one (representing a score of 4) checked.
- How much does the answer influence your next actions?:** A 5-point scale from 'Not at all' to 'My next actions depend on the answer'. The scale has five radio buttons, with the first one (representing a score of 1) checked.

Figure 1: Example diary page (completed)

Main User study.

The user study explored which questions people would like to ask on mobile devices. In particular, we were interested in the relation of the question to user’s location and the degree to which the answers may influence the next user activity. Participants were given a paper diary as shown in Figure 1.

Participants were asked to record their (mobile) information needs by noting the question (see top part of the diary page) and the location they were at when they wished to ask the question (second line in diary page). In addition, we aimed to find out how much the question related to the place. The motivation was to find out if knowing the place one may predict questions, or if people at the same place may find information about other people’s questions insightful. We gave a scale to describe the degree the question was related to the place (from 0 – not at all, to 4 – very strongly). For example, a question may have been directly inspired by the place (e.g., a museum’s exhibit), or it may be independent of the place and the user’s surroundings. As a result from the pilot study, we also gave a second scale for the degree to which the participant’s question related to their next activity. Both scales are given in the lower part of the diary page. Each page in the diary referred to one question.

4. RESULTS

The participants were computer literate and familiar with mobile electronic devices. 12 people participated in the diary study, which was recorded over one week. A total of 220 unique entries were recorded, with an average of 18 entries per participant. The minimum number of entries in the diary was 8 and the maximum was 29. We analysed a number of different aspects for the diary entries, which are now discussed in turn.

Relationship to Place (Query/Location).

We first explore the relation between question and location by

analysing the user location and the references to locations in the query itself. We identified five categories of places: the participant's home, a friend's home, their work place, going out (e.g., shopping, dentist, airport), and being mobile in the road (e.g., in a car). Figure 2 shows the number of queries per category. We observe that almost half of all queries were asked at home; all other categories are locations away-from-home. Work and friends' places are similar to home as the participants were mostly indoors. The low number of queries of mobile participants most likely reflects the difficulty of recording questions in a diary while driving.

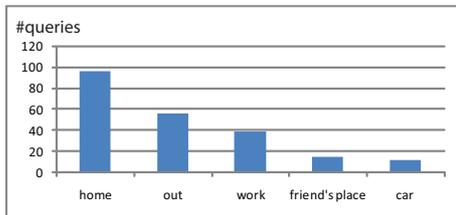


Figure 2: Query location by categories

The actual places of querying (i.e., location of the participant while querying) and the location scope of the query (i.e., location the answer refers to) are shown in Figure 3. For example, participant P10 recorded the query “Where is Koh Samui?” in Hamilton, but the scope of the query is Thailand.

The location of the user's querying was taken from the user statement in the diary. The scope of the queries was determined by analysing the query text in the diary. It is not surprising to find that most queries were issued in Hamilton ('home' city of the study); the other three locations of queries are within 200km from Hamilton.

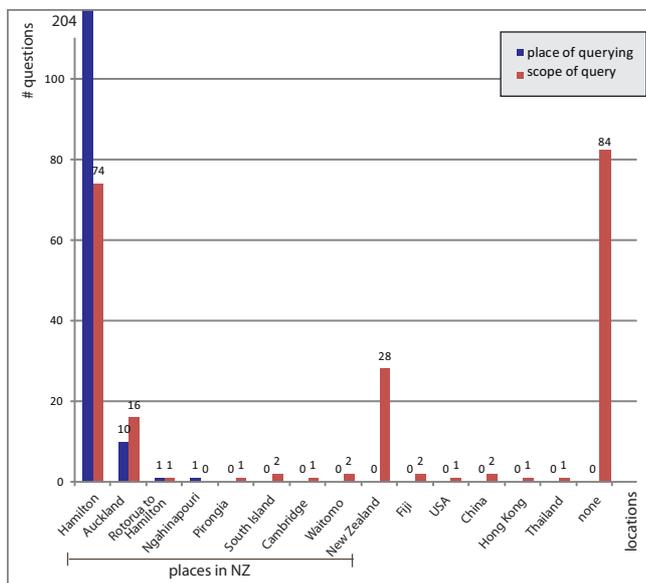


Figure 3: Location of querying vs scope of queries

The location scope of the query was deduced from the queries: by place names explicitly mentioned (e.g., “Hong Kong”, “South Island”) or by landmarks mentioned (e.g., “Waitomo Caves”). The more general scope of ‘New Zealand’ was inferred from the query context (e.g., “What is the song [currently] playing on the radio?”).

Factual queries mostly have no location reference (e.g., “What is the best chess tactic?”). From our analysis of the query text of the 220 queries, more than a third (80, 37%) have no location reference; 47 (21 %) had implicit place references (“how to predict the interest rate in the next month” is assumed to refer to a NZ context; “which restaurant is best” refers to a Hamilton context).

Question Type (question words used).

We first analyse and distinguish questions submitted by the participants by the *question words* they used. Question words are the ones starting the question (e.g., ‘how’, ‘when’, ‘where’), a small number of which were corrected for grammar. The participants in our study gave complete questions, not just keywords (as common in web search). Diary studies enable a focus on the participants' questions instead of a collection of system-influenced query terms. Paper diaries in particular do not suffer from input mechanisms influenced by the recording device.

Figure 4 shows the participants' question words used at different places. Most questions are asked at home: one reason may be the ease of use of the diary at home. We immediately see that the two main questions asked at home are of the form “what is this?” (e.g., P1: “What is the song playing on radio?” and “How to do this?” (e.g., P9: “How can I improve my thesis?”), asking for some kind of information and advice. Note that the ‘what’ questions still exist at the workplace, but the ‘how to’ questions are diminished. In this analysis, we further split the ‘out’ category into ‘public place’ and ‘shopping place’ to highlight an observation: If taken together in an ‘out’ category, the ‘how’ and ‘where’ questions occur in almost the same number. However, using our split category, we see that when shopping, the participants mainly ask ‘how’ questions and in other public places they use ‘where’ questions. Analysing the data, we find that queries in shopping places often refer to pricing and selection of goods (e.g., “How much does a Gibson guitar cost?”); whereas questions in public places mostly refer to geographical features (e.g., “Where is Pumice Cafe?”).

Question Type (categories).

We identified two main categories of questions: problem solving (*problem*) and geographic references (*geographical*); for each we found a number of sub-categories, some of which were again subdivided (see Table in Figure 5). Our categories are related to those used in other studies [4, 1], but is chiefly influenced by the available user data.

Figure 6 shows how the question words relate to the categories we identified. We see that there is no simple mapping that related each question word to one meaning only. The figure also shows the distribution of question words used. The occurrence of some questions words (e.g., ‘why’ and ‘will’) is so low that it is hard to draw conclusions about their contact of usage.

Using categories and subcategories, Figure 7 shows an overview of the questions asked by each of the 12 participants. We see, e.g., that participant P8 submitted overall 10 questions in two categories only (information and direction), and that participant P9 mostly asked for advice. Comparing the participants, we see personal preferences but no overall pattern regarding the categories used emerges yet at this level.

This changes, once we compare question categories at different locations as shown in Figure 8. We can identify the following patterns: Problem-solving questions (asking for information, advice, or facts) are mainly asked at home. The only other place where fact-related questions are asked is at work; these questions occur in no other places. People express directional information needs (di-

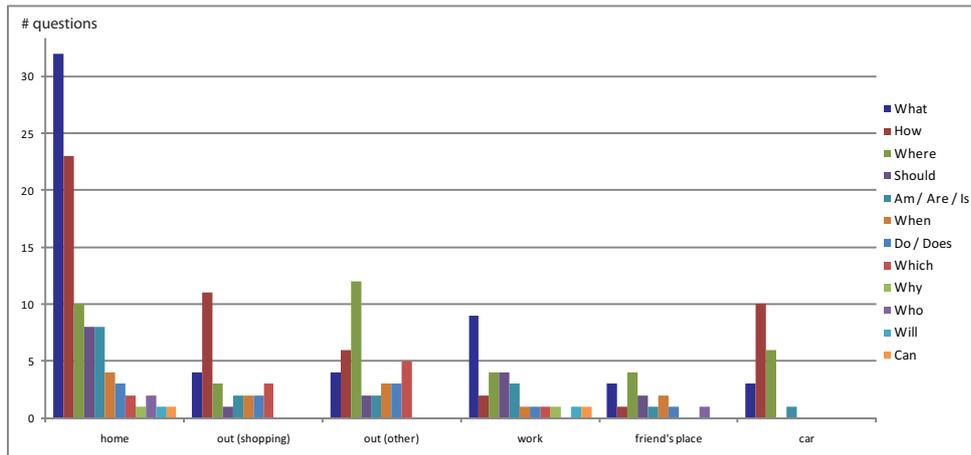


Figure 4: Questions words and locations

Category	Sub-category	Explanation
Problem	Information	The answer is published by an authority or organization (e.g., bus timetable)
	Fact	There is only one answer for question (e.g., 2+2)
	Advice	The user is looking for reports of experiences others made (e.g., how to ...)
	Food	... concerning food or cooking
	Shopping	... concerning spending money on something
	Activity	... concerning an activity
	People	... concerning people
	Place	... concerning places
	Traffic	... concerning traffic
	PP	... concerning personal problems
	Health	... concerning health
	Beauty	...concerning skin care etc
	Skill	...concerning improvements of skills
	Gift	...concerning gift suggestions
Entertainment	...concerning movies, trivals or books	
Geographical	Direction	The answer is a direction
	Place	...to a particular place
	Close by	...show/name all places near-by
	Time	The answer is an estimate of time to reach a place
	DA	Answer is direction combined with advice (e.g., what to do while waiting for tyre to be fixed)
	Shortcut	Answer is a description of a quick way to get to a place.

Figure 5: Categories of Questions

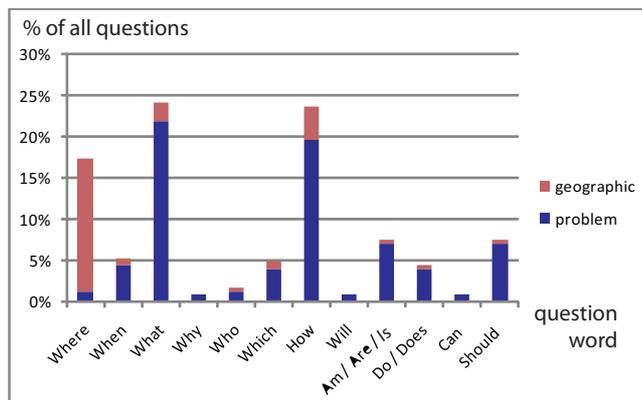


Figure 6: Relationship between question words and identified question type

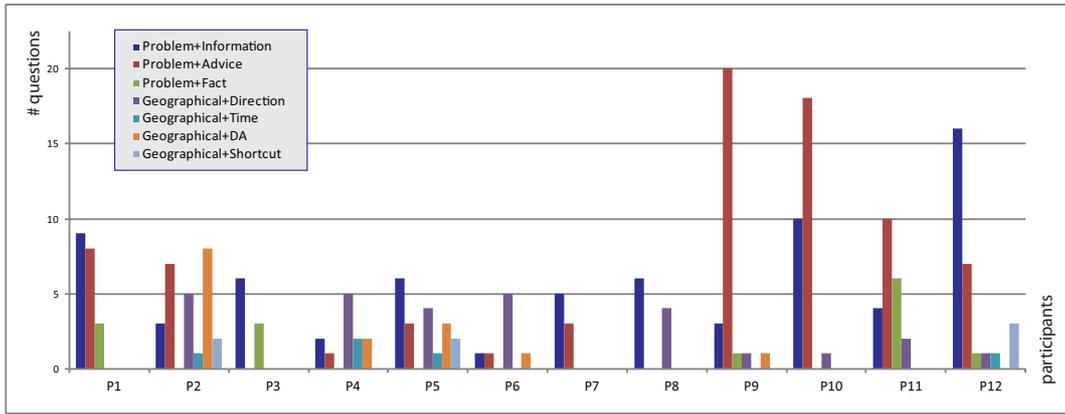


Figure 7: Question type over participants

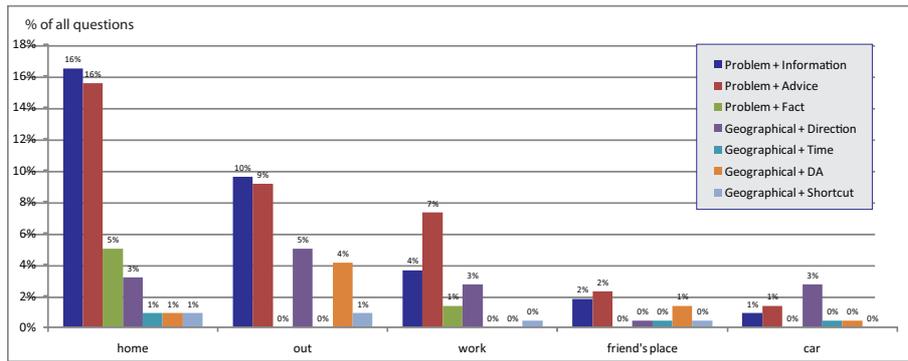


Figure 8: Question type over location of querying (categories)

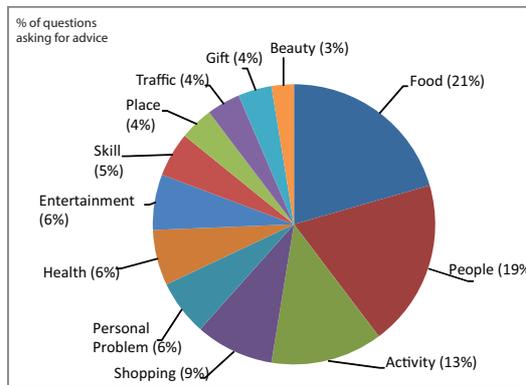


Figure 9: Topics occurring in questions asking for advice

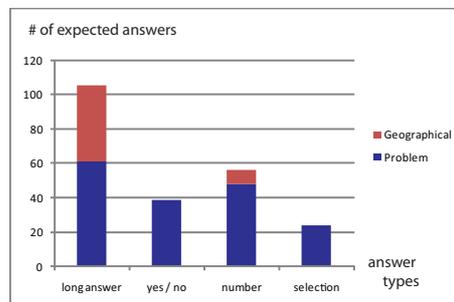


Figure 10: Distribution of answer types

rection, DA and shortcut) mostly when they are out (10%) followed by home (5%), work and mobile (3% each).

Geographical questions are the prevalent type of questions when being mobile – the number would probably have been higher when using different recording mechanism for questions (e.g., dictaphone). Direction and advice is almost only found when people are out. They are typically of the form of “here is my problem, where do I go for a solution?” such as “Where can I buy a number for my mailbox?” (P4).

So far we only used categories and sub-categories in our analysis. As shown in Figure 5, the sub-categories of Advice and Direction have further distinctions.

Figure 9 shows the distribution for advice-questions: questions about food (e.g., “What can I cook for dinner tonight?”), people (e.g., “Which of my friends plays any music[al] instrument?”) and activity (e.g., “When will the roses need to be pruned again?”) were most common. Amongst the direction questions, 64% referred to places known to the participant (e.g., “Where is [the] Genesis office?”), whereas 36% asked for information about the location of nearby-by places (e.g., “Where is the nearest Garage?”, “Where is the nearest petrol station?”).

Expected Answers.

We analysed the questions for the expected length of the answers. Figure 10 shows the distribution of long answers, simple yes/no decisions, numbers and selection for the given set of queries. Note that answers for problem-solving questions and geographical questions may have very different structures: For example, the long answer to a geographical question may be a map rather than a textual document as for the problem-solving question. Numerical answers to geographical questions typically answer questions for estimated travel times (e.g., P4: “How long it take to drive to Auckland airport?”).

Relationship to place and activity (numerical score).

We now discuss the numerical aspects of the study relating to the two scales shown in the lower part of the diary page (see Figure 1). We identify four groups of questions: independent of location and activity (score 0 on both scales), relating to activity only, relating to location only, and relating to both activity and location.

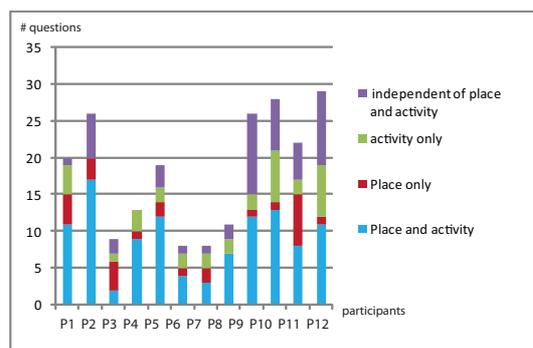


Figure 11: Relation between questions and the user's location (place) and current/next activity

Figure 11 shows the number of questions for each participant and their relation to place and activity. 109 out of 220 questions (50%, avg 9) referred to both location and activity; ranging from a minimum of 2 to a maximum of 13. This is the largest category in diary study. 34 queries relate to activity only (16% avg 3, max

7, min 0); 27 to location only (12 %, avg 2, max 7, min 0). 49 questions were independent (22%). Examples are: “What is Type II diabetes?”, and “[How much do] you have pay for on over-speed ticket?”.

On average, 22% of all questions did not relate to place nor activity. The number of independent questions per participant ranges from zero to 11. Note that the strength of the relationship (scores between 1 and 4) cannot be seen in Figure 11.

Scores between 1 and 4 describe how strongly the question relates to (a) the place and (b) how much the answer influences the next action (zero if independent of place/activity). The scores for place and activity relation are shown in the two diagrams in Figure 12. For clarity, independent queries (place and activity relationship zero) are not shown in this figure.

The average score on location (place) relationship for is 3.55 for queries that relate to both location and activity (88%), and 2.93 (73%) for those relating to place only. The average score for relationship to place is 3.53 (88%) for queries that relate to both location and activity, and 3.03 (75%) for those relating to activity only. We observe that if a question relates to both location and activity, this relationship is experienced as being stronger than if the question relates to only one of them.

Figure 13 shows the relationship between the question word used in the queries and the average scores for location and activity. We observe that ‘who’ and ‘can’ are not used for questions relating to the next activity. ‘Where’-questions often influence the next activity (as in, “Where is the nearest xyz?”), whereas ‘which’-questions relate more strongly to the location (often occurring in questions like “Which is the quickest way to place xyz?”).

Places and location/activity scores.

Figures 14 and 15 show the relationship between the scores and the place of querying and the scope query, respectively. In Figure 14, we see that participants deemed the relationship between their activity and location to the question to be strongest when being mobile in a car. This matches the observation that questions for participants in a car were mostly geographical (often queries for directions relating to their immediate surroundings). Again, we'd like to stress the observation that the data does not necessarily mean that people do not have any other information needs while driving, but that they did not record them.

We observe that both at their own ‘home’ and at ‘friend’s homes’, participants asked questions more strongly related to their next activity than to their location. These two places also show the lowest overall scored for relation to place. One explanation could be that these places are seen as ‘default’ locations from which to start other activities; whereas work, going to the city, and driving, are seen as more explicit activities that may create their own questions. We also observe that there does not seem to be a dependency on how long/often people spent time in these places.

A number of observations can be made about the comparison between scores and location scope of the query (Figure 15). Firstly, we observe that the average scores in this Figure are not comparable as some of them represent single queries (indicated by gray background) and others more than 70 queries (as for Hamilton, or ‘none’). We can, however, see that individual queries mentioning places outside New Zealand are interpreted by their authors very differently: referring strongly to activity (as in the case of Thailand) or not at all (e.g., USA). For some queries referring to places in New Zealand (e.g., “Rotorua to Hamilton” and “Cambridge [New Zealand]”) are related to the user's location. In the first example, the participant P8 was on the way from Rotorua to Hamilton, asking for the nearest petrol station on the way. In the

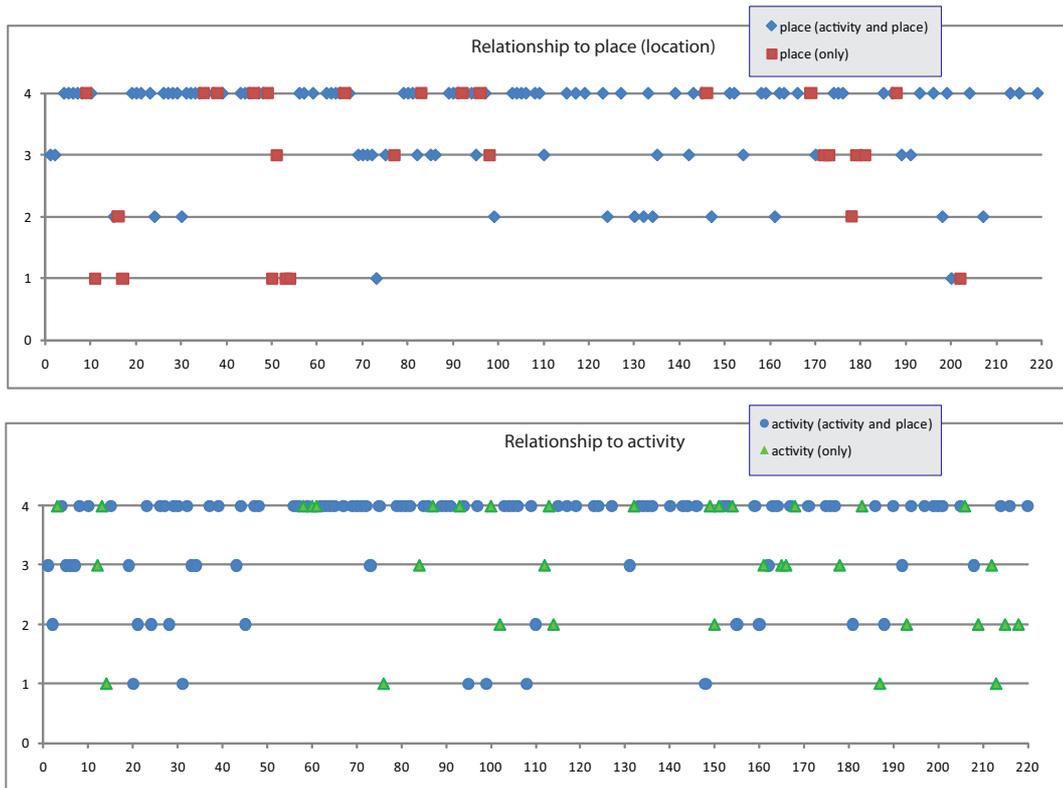


Figure 12: Scores for queries relating to place (top) and activity (bottom)

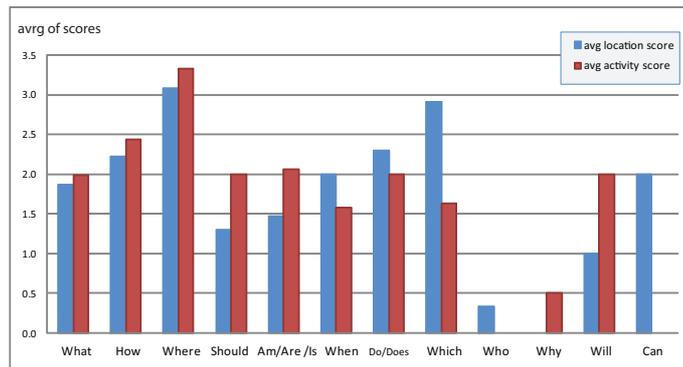


Figure 13: Relationship between question words and place score

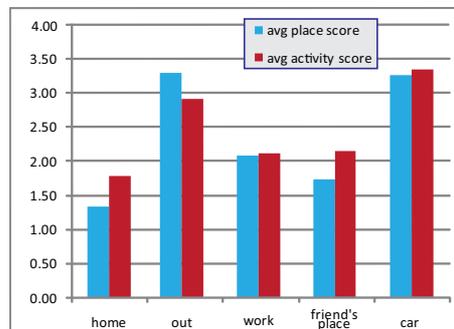


Figure 14: Relationship between activity and place score and the user location when querying

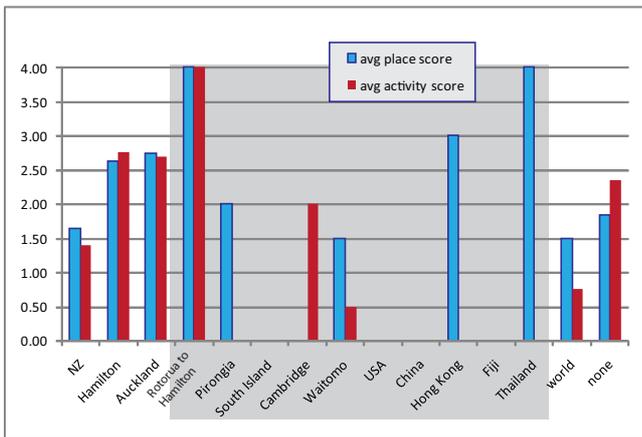


Figure 15: Relationship between activity and place score and the answer's location (determined by users)

second example, participant P12 was asking for the way from their home to Cambridge (with the intention to go there).

Some of the scores for queries in places categories such as "New Zealand", "world", and "none" appeared to contradict their query text. We believe that these three categories have only very loose relationships to place. Examples are "Can you get puzzle quest on PS2?" (P12) with a NZ scope (for being able to buy the game in the country) and location and activity scores of zero. In this case, the views of participant and authors align. However, for a large number of cases, they do not (as shown in the high scores for location-relevance in these three place categories). Examples of such diverting views are P9: "Do I have to send any emails?" and P10: What is on TV tonight?, which were both given very high scores for location by the participants, however, no relationship to place could be inferred from the question alone (query location scope being 'none' and 'NZ').

5. DISCUSSION

We first compare our findings with those of earlier studies. Kamvar & Baluja expressed the desired to determine "what inspired [a] search" – our recording of query place and its relationship with place and activity give indication to those user motivations. However, we observed discrepancies between the participants judgement of strength of relationships and the results of our query analysis, which are hard to resolve without further data.

We observed an influence of location and activity on the query in 78% of all questions, which is comparable to Sohn et al.'s findings regarding context [17]. The proportion of queries asked in the car is low (6% of all queries), but 62% of these ask for directions or estimations of driving time. The lowest relative number of geographical questions were asked at home (14% and at work 21%). However, *explicit* references to locations (perhaps comparable to Church & Smyth's 'geographical terms' [4]) were found in only 69 questions (31%). 52% of those occurred in geographical questions, and 48% in problem-solving ones. The reason may be that participants felt that their location context was somehow implicitly known. We observe that counting the explicit location references does not give a good indication of the type of the question. The question words used (as shown in Figure 6) provide a better indication but are not typically used in online searches, where log analysis studies indicate shorter average query length (2.56 words [10]). Explicit location references for users in the car were mentioned in 6 of

13 questions (46%), which is consistent with the greater frequency of geographical references reported in [4].

The (implicitly assumed) location context of users is not always easy to measure. Location coordinates are not sufficient, as places and areas need to be identified. The opposite problem is easily solved (finding locations for a given place or building); identifying the context the user sees themselves in is not that simple. Given coordinates may indicate anything between 'Bongo cafe', 'University of Waikato', 'North Island', 'South Pacific', and 'Planet Earth'. Deciding which of these nested location contexts is appropriate, requires further work [15].

We can confirm the observation that most queries were made when being "off-duty"[12] (assuming that we can infer the participants' status from from their location outside work). However, the strong concentration on public place could not be confirmed (only 28% of all queries were issued in a public place). The influence of social interaction was not explicitly addressed in our study. However, 20% of the questions explicitly mention other people or their places ("our visitor", "my friend", "Lesley" telephone number", "Sam's place").

Amin et al[1] reported that in searches within the participants' vicinity, target locations were more often those regularly visited (that is, participants engaging in long-term planning). However, we discovered that most searches for places near-by were performed when being on the road (that is, an immediate information need is addressed). This observation is confirmed by an analysis of the queries issued: of the 10 searches explicitly mentioning "near-by" context, three referred to finding car parks and two to finding petrol stations (others: shops (2), restaurants (2), hospital (1)). Thus 50% of vicinity searches were issued in the car, all of which referred to car-related immediate issues.

Although there is a sizable minority of queries that are independent of location or activity, we found these factors had strong influence on participants queries. This is consistent with earlier studies: similar to [17], we found that 78% of queries were prompted by the participants' situation. However, we found that 28% of the queries reported were seeking details about the present location of the user (compared to 15% by Nylander et al. [14]).

We analysed the queries for sources of possible answers (see Figure 16, top). The topics of 150 of the 220 questions (68%) were such that the answers, in our judgement, are likely to be available on the Internet. This includes queries that would have to be reformulated or extended by (currently implicit) context information. For a subset of 68 questions (31%), which did not contain references to current events (such as exhibitions), the answers may be obtainable from a digital library. Given that some questions may be answered by both, Internet and library, this leaves 63 questions (29%) that may not be answerable by such existing online services. A large number requested personal assistance in decision making such as "Should I pick up the pizza or use delivery?". However, the majority of those queries contained references to user's immediate surrounding, e.g., "Does this place sell dried peas?". However, these observations need strengthening by evaluation using actual online services.

We evaluated the completeness of the queries (see middle of Figure 16): 22 questions were ambiguous and 17 were under-specified. Ambiguous questions were deemed to be those it was not clear to the authors what an expected answer should be, for example: "Should driving age been raised to 18 years old?" (reflecting a current public discussion in New Zealand) and "Should I make a coffee?". Under-specified questions were those that needed further information to be answerable that could not be automatically detected or observed (such as location would). Examples were "What is the

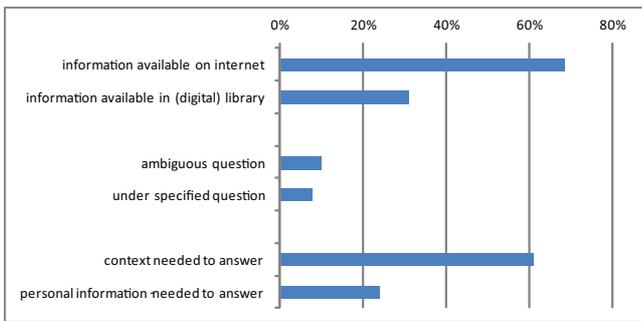


Figure 16: Quality/suitability of queries

postfee on this parcel (milk powder) to china?", "Do they have a book I read about recently?" Amongst those with explicit location references, a number were under-specified (e.g., "the New World shopping centre", "where should we meet at St Paul?", "does *this place* sell dried peas?" – further information was needed to clarify which place was referred to). Under-specification mostly referred to location, but other instances of known context were also assumed, e.g., "Where else can we go have a look for vanity and bathroom applications?" assumes knowledge of the places that have been visited (history and shop identification), and "The best size of bath tub that can fit well to the main bathroom?" "Which is the best size of bath tub can fit well to the main bathroom?" assumes knowledge of the participant's house layout; and "Is my server running?" assumes information about and access to a private server.

A large number of questions needed additional context information to be answered (see bottom of Figure 16). Not surprising for mobile queries, most of these contextual questions required information about the user location. Location context can be either direct coordinates (to answer all questions like "where is the nearest xyz?" and "How do I get to xyz from here?"), or information about the city and country (e.g., "How will tomorrow's weather be like?" needs a place reference, and P1: "How many people have swine flu in Hamilton?" needs a country reference). Some questions require a personal scope as location reference, such as "Where can I catch the first glance of the sun on next Friday?". Other contexts needed may be locations of shops or other people's homes. The context is often temporal, not only asking for the current location of the active user, but also questions that require knowledge about the current locations of other (mobile) people or items (e.g., "Where should I meet my friend in St. Paul? ?" and "where is my mouthgard?"). Surprisingly, there were no questions to locate other people, only items and shops ("Does this place sell dried peas? Where are they?"). The previous query also shows an example of needing to identify places by more than their location, that is, to identify their semantics (mostly as shops). References to places were: "they", "this place", "here", "this car park building", "this shop".

Identification of persons is also part of context needed. Other people were often simply referred to by name, especially friends (e.g. in "Carole's place"), which could be resolved by the system. Sometimes, however, people were referred to by indirect reference – as if they were pointing at the person (e.g. "what is this boy's name?"). One query is particularly interesting in this respect: The query was "Is PolyMike leaving Uni?" (name changed to preserve anonymity), requiring knowledge about a person's nickname. One way to obtain this context information could be a search in a list of personal friends, or an online search in personal home-pages held by staff members to map the nickname onto the person's name to identify their location.

The definition of the categories and question types we used were driven by our research questions and the available data. They are similar to the ones used in other studies [11, 14, 17, 4, 1, 6], without attempting an exact match. We observe that some of the categories find an overlap, e.g., food, shopping, travel, whereas some significant categories found in other studies did rarely occur, e.g., 'news' and '(public) transport'.

We confirm the finding that a significant number of queries are context free, or appear to the participants to be so: 22% of queries were marked as being independent of location and activity. As noted earlier, some discrepancies were observed between user and author views. Further study is needed.

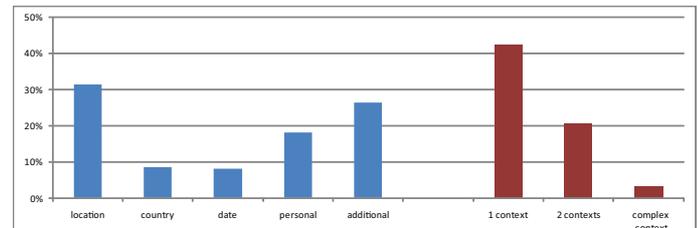


Figure 17: Context needed

Figure 17 lists types of typical context information required to answer the questions. 31% of all queries need at least information about the user's location to be answered – the best example being (P8): "Where am I?", and the whole range of requests to identify the nearest petrol station, super market, car park etc. Sometimes the reference to the city or even only the country would be sufficient: typically here the users implicitly assumed a New Zealand context without stating so in the question (e.g., P1: "How many people have swine flu in Hamilton?"). In all cases, the missing information could be obtained from the current position of the user, that is, participants assumed their current context to automatically hold for the query. Some queries require the current date/time to be answered, e.g., when asking for a song playing on the radio, movies being on at a cinema, the next bus to arrive, or estimated travel times by car.

18% of all queries required information about personal data to be answered. Examples as questions about finances (P1, P10, P11, P12) or emails (P9) to be answered, but also information about friends and acquaintances (P6: "Which of my friends play a musical instrument?", P11: "What is Lesley's telephone number?", and P11: "Has PolyMike left the uni?"), and personal advice ("Should I do cardio or weights?", "What should I text my friend?").

Jones & Buchanan propose a Questions-not-Answers system that provides users with information about the questions other people asked in the vicinity of the current user [8, 2]. Our study results show that this approach is not valid for all questions, users may ask given the significant proportion of independent questions. Additionally, geographical and problem-solving questions may have to be treated differently. For geographical questions, providing the location of the answers would be more meaningful. As an illustration, it is not hard to imagine a spot with unhelpful or missing road signs on the way to a favourite tourist location. If one happens to get lost at that point and wants to query for the location of the tourist site, seeing all other queries asking for that very site may not just be unhelpful but also misleading. However, one would hope that for the government body responsible for road signage, this information may be of high interest. The research reported here does not yet answer the question of when and where to display information about questions or answers, but explores the questions of the relationship

between a query and the location of the user, and the relationship between query and the next action of the user.

Although there are some differences in methodology, between this study and other diary studies of mobile information behaviour [17, 4, 1, 14], findings on the importance of context appear to be robust. Our study gives some suggestions that some study parameters appear not to significantly alter the results, including: the medium of the diary (paper or device), recording technology (pen or text input) and whether participants are explicitly reminded to record entries. However, we still believe that future similar studies would be valuable to confirm these methodological aspects and to further examine differences in participant background and device familiarity. In particular, we suggest that further studies utilising mixed recording methods [12], that allow participants flexibility in how they record entries, may help to improve validity. For example, Palen and Salzman suggest voice-mail diaries [16] whereas Brandt et al. propose txt messaging snippets for later clarification [3]. Finally, it is worth noting that although we have attempted to study spontaneous information needs 'in the wild', it is of course the case that the presence of a new device with enhanced capabilities will necessarily change the users' environment and their behaviour.

6. CONCLUSION

In this paper, we reported on the results of a diary study with mobile users. Participants recorded their information needs, the place and date, and how strongly their question was related to their location and next activity. We found that the questions asked by mobile users are strongly influenced by contextual factors; however, a significant portion of independent questions (without reference to location or activity) was found.

We found that the type of questions recorded by participants varies across their locations (home, work, at friends homes, out in the city and while driving). These variations occur both in the query terms and in the form of desired answers.

We observed that for detecting context (e.g., to identify location-related questions), analysis of query keywords alone is not sufficient. More information was gained by analysing the query words (e.g., when, how, what) – which requires the queries to be given as full questions (instead of keywords). A number of pre-processing steps would have been necessary to answer the user questions: correction of spelling; semantic resolution of references to people's names and identification of places ("home"); as well as detection of (currently) implicit user context (e.g., location, date, people-near-by). For these steps, additional (personalised) services, such as PIMs and gazetteers may be employed.

Both the location of queries and the participants' activities affected participants' questions: when information needs were affected by both location and activity, they tended to be strongly affected by both factors.

Overall, we found that a number of interacting contextual influences shape a user's mobile information needs. None of the categories, topics, and other characteristics used to describe the user queries are strictly orthogonal, and none of them can be used in isolation to identify the users' information needs or context. Thus adaptation of devices and interfaces need to take into account a rich variety of situational factors.

7. REFERENCES

- [1] A. Amin, S. Townsend, J. Ossenbruggen, and L. Hardman. Fancy a drink in canary wharf?: A user study on location-based mobile search. In *IFIP Human-Computer Interaction (INTERACT '09)*, pages 736–749, 2009.
- [2] D. Arter, G. Buchanan, M. Jones, and R. Harper. Incidental information and mobile search. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 413–420, New York, NY, USA, 2007. ACM.
- [3] J. Brandt, N. Weiss, and S. R. Klemmer. txt 4 l8r: lowering the burden for diary studies under mobile conditions. In *CHI'2007: Extended Abstracts, Conference on Human Factors in Computing Systems*, pages 2303–2308, 2007.
- [4] K. Church and B. Smyth. Understanding the intent behind mobile information needs. In *IUI'09: Intelligent User Interfaces*, pages 247–256, 2009.
- [5] Y. Cui and V. Roto. How people use the web on mobile devices. In *WWW'08: World Wide Web Conference*, pages 905–914, 2008.
- [6] D. Dearman, M. Kellar, and K. N. Truong. An examination of daily information needs and sharing opportunities. In *CSCW'08: ACM Computer Supported Cooperative Work*, pages 679–688, 2008.
- [7] T. Heimonen. Information needs and practices of active mobile internet users. In *Procs. of the 6th international Conference on Mobile Technology, Application & Systems (Mobility '09)*, pages 1–8, 2009.
- [8] M. Jones, G. Buchanan, R. Harper, and P.-L. Xech. Questions not answers: a novel mobile search technique. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 155–158, New York, NY, USA, 2007. ACM.
- [9] E. Kaasinen. User needs for location-aware mobile services. *Personal Ubiquitous Comput.*, 7(1):70–79, 2003.
- [10] M. Kamvar and S. Baluja. Deciphering trends in mobile search. *Computer*, 40(8):58–62, 2007.
- [11] M. Kamvar, M. Kellar, R. Patel, and Y. Xu. Computers and iphones and mobile phones, oh my! In *WWW'09: Int. World Wide Web Conference*, Apr. 2009.
- [12] I. Lee and J. Kim. Use contexts for the mobile internet: A longitudinal study monitoring actual use of mobile internet services. *International Journal of Human-Computer Interaction*, 18(3):269–292, 2005.
- [13] T. Matthews, J. Pierce, and J. Tang. No smart phone is an island: the impact of places, situations, and other devices on smart phone use. Research Report RJ10452, IBM, Sept. 2009.
- [14] S. Nylander, T. Lundquist, A. Brännström, and B. Karlson. "it's just easier with the phone" — a diary study of internet access from cell phones. In *Pervasive'09: Int. Conf. on Pervasive Computing*, pages 354–371, 2009.
- [15] W. Osborn and A. Hinze. Issues in location-based indexing for co-operating mobile information systems. In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, pages 226–235, 2007.
- [16] L. Palen and M. C. Salzman. Voice-mail diary studies for naturalistic data capture under mobile conditions. In *CSCW*, pages 87–95, 2002.
- [17] T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan. A diary study of mobile information needs. In *CHI'08: SIGCHI Human factors in Computing Systems*, pages 433–442, 2008.
- [18] H. Wigelius and H. Väättäjä. Dimensions of context affecting user experience in mobile work. In *INTERACT '09: Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction*, pages 604–617, Berlin, Heidelberg, 2009. Springer-Verlag.

© ACM, 2010. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in MobileHCI '10, <http://doi.acm.org/10.1145/1378889.1378924>