Entertainment on the Go: Finding Things to Do and See while Visiting Distributed Events

Richard Schaller
Computer Science (i8)
Uni of Erlangen-Nuremberg
Haberstr. 2
91058 Erlangen, Germany
richard.schaller@cs.fau.de

Morgan Harvey
Computer Science (i8)
Uni of Erlangen-Nuremberg
Haberstr. 2
91058 Erlangen, Germany
morgan.harvey@cs.fau.de

David Elsweiler
Institute for Information and
Media, Language and Culture
University of Regensburg
93053 Regensburg, Germany
david@elsweiler.co.uk

ABSTRACT

Distributed events are collections of single events taking place within a small geographical area at approximately the same time, normally related to one given topic e.g. music, film, arts etc.. There are usually a large number of events on offer and the times in which they can be visited are heavily constrained. Therefore the information seeking task of choosing the events to visit and in which order can be very difficult.

In this paper we investigate, via 2 large-scale naturalistic studies (n=391 and n=740), how mobile applications can be designed to assist users in this task and how such applications are used. We present an application that allows users to search and browse the events on offer in a number of different ways including via personalised event recommendations. Logs were collected of user interactions with the system. The results of this log analysis in combination with 2 surveys show some surprising usage patterns and point to how such applications can better serve users' needs.

Categories and Subject Descriptors

H.5.2 [**Information Storage and Retrieval**]: User Interfaces—*User-centered design,Evaluation/methodology*

General Terms

Design, Experimentation, Human Factors

Keywords

Mobile Assistance System, Distributed Events, Information Needs, Mobile Search, Mobile Information Seeking, Casual Leisure Search

1. INTRODUCTION AND MOTIVATION

A distributed event is a collection of smaller, single events occurring at approximately the same time and conforming to one overarching theme. Well known examples include the Cannes International Film Festival, the Edinburgh Festival Fringe, and Montreal International Jazz Festival. What many of these events have in common is that they have huge number of diverse sub-events that are geographically and temporally dispersed. For example, in 2011 the Edinburgh Fringe had a total of over 2,500 shows at 258 venues all over the city, ranging from the classics of ancient Greece, William Shakespeare and Samuel Beckett to modern works.

While variety is the biggest selling point of these events, making the decision of which sub-events to visit will not be straight forward. Visitors will have to discover which events are on offer and select a small subset from a list of typically hundreds that they find appealing. The decision will moreover depend on factors such as user preferences, time constraints, the location of the event and perhaps transport connections. Often, as a result of the sheer choice, people feel overloaded with information and rely on tips from friends and on serendipitous discovery [22]. Without careful planning this could mean missing events that would be of interest or spending large amounts of time travelling between locations. Information systems, such as mobile phone apps [22], web based portals e.g. [30] or recommender systems [16] present a clear opportunity to assist users in performing this difficult task.

This is a topic of interest to the interactive IR community as it is an example of a non-work task; an area where we have only a very limited knowledge of how people behave to resolve needs, and how systems should be designed to best support users achieve this [10]. In this work we examine, based on 2 large-scale studies, how visitors to distributed events can be supported by information systems to maximise their enjoyment of the event. The main contributions of this paper are as follows:

- We perform interviews to establish user needs / priorities in the context of two specific distributed events
- We present a system with features designed to meet some of these needs
- We examine usage of this system with two large scale studies
- We report on how different features of the system were used and how this influenced events selected
- We try to infer how different features supported the needs of the user as reported in the interviews

IIiX 2012, Nijmegen, The Netherlands

2. RELATED WORK

The work in this paper straddles 3 areas of related work. Section 2.1 summarises related research from the information seeking community; Section 2.2 deals with mobile search interfaces; and Section 2.3 outlines work on systems designed to assist tourists.

2.1 Leisure Information Seeking

Information seeking behaviour is traditionally studied in the context of people completing work tasks. Despite its name, a work task need not be work-related. It is simply a sequence of activities a person has to perform in order to accomplish a goal [15]. A work task has a recognisable beginning and end, it may consist of a series of sub-tasks, and results in a meaningful product [3]. Correspondingly, the models we have of information seeking behaviour tend to assume that people look for information in response to a lack of understanding [6] or the recognition of a gap in knowledge [2] preventing the completion of the task at hand.

Based on two investigative studies, one examining information needs in the context of television viewing and the other analysing broader information behaviour reported on twitter, Elsweiler and colleagues [10] proposed a model for what they refer to as casual leisure search, which deviates from standard work-based models. According to their model, in casual-leisure situations, users seek information not in response to a knowledge gap, but with the aim of being entertained or passing time. Such needs tend to be directly related to mood, physical state or the surrounding social context. A further defining characteristic of such needs is that the informational content found by users is often less important than the feelings induced by the found content and/or the search process itself.

Beyond these two studies, very little literature explicitly focuses on information seeking behaviour in casual-leisure situations. Exceptions include studies of finding fiction [20, 21] and non-goal oriented newspaper reading [29]. As a research community we have very limited knowledge regarding how people respond to meet casual-leisure information needs or the kind of support that should be provided to help users achieve their aims.

2.2 Mobile interfaces

Previous research on mobile information seeking has primarily focused on web search conducted on the go. Kamvar et al. investigated search logs collected from both desktop and mobile versions of Google [18]. They found that while mobile search differs from desktop search in that queries are often much shorter, this is not the case with more modern smartphone devices. On a per-session basis it was found that mobile users tend to submit a smaller number of queries and tend to display less topical diversity in their searches. In more recent work Teevan et al. [28] present a survey of local mobile search demonstrating that both location and time are crucial contextual factors affecting search behaviour, not just in terms of current location but also in terms of items of interest on the route and at the user's intended destination. They conclude that an understanding of location, time, and social context could lead to an improved search experience for users.

Church et. al investigate [5] the intent behind information needs in a mobile context and also observe that location and time are crucial factors. More recent work [4] evaluates mobile user interfaces and shows that consideration of the interface has a significant impact on how users interact with the system. The authors suggest that this choice should depend on various factors including a user's personal preferences, their information need and their situational context. This means that a single interaction paradigm may not optimally suit all users' needs and therefore a choice of interfaces allowing different methods of access to the same content may be preferable.

2.3 Tourist Assistance Systems

There is also a body of work within the HCI and AI communities that focuses on developing Tourist Systems to assist visitors in discovering and getting to points of interest, usually within a single city. Both of these assistance problems are also relevant to distributed events. In the literature several approaches for selecting POIs are reported. The most basic approach is a system that allows the user to manually select POIs from a list [19]. If the system is used on the go then the number of POIs can be restricted by only taking into account nearby POIs or those which are currently available or open [11].

Most tourist systems utilise a content-based recommender based on POI metadata and user profiles [23] or collaborative filtering based on other users' preferences [11]. To further assist users in finding the selected POIs systems display them on a map [11] or calculate a suitable footpath to them. More elaborate systems calculate an optimal tour of the selected POIs [19]. Automatically deciding which POIs should be visited makes it possible to optimise both the visitor's preference for specific POI types and the length of the generated tour [14].

We build on all of the literature described above in our work. Our system combines aspects of the described search interfaces with other modes of interaction to achieve the flexibility endorsed by the Church et al.'s work. The proposed system also builds on the state of the art in Tourist Systems by intelligently combining chosen events into a plan and guiding the user through each step in the plan. Finally, we contribute to the information seeking literature by learning how people behave in a specific casual-leisure context.

3. THE LONG NIGHTS

In this paper we focus our efforts on two particular distributed events. These are the Long Night of Munich Museums (Lange Nacht der Münchner Museen, LNMuseum), an annual cultural event organised in the city of Munich, Germany[1] and the Long Night of Science (Lange Nacht der Wissenschaften, LNScience) in the metropolitan area of Nuremberg, Erlangen and Fürth[2]. During the Museums Night a diverse range of small and large museums, as well as other cultural venues, such as the Hofbräuhaus and the botanical garden open their doors during one evening in October. Participants at the Science Night include companies, Universities, Colleges and other public facilities who use the event to demonstrate their research and allow the general public to understand their work. On both nights many venues organise special activities and exhibitions not otherwise available.

[1]The event is organised by Münchner Kultur GmbH (http://www.muenchner.de/museumsnacht/)

[2]The event is organised by Kulturidee GmbH (http://www.nacht-der-wissenschaften.de/)

Visitors to these Long Nights include both locals and tourists and represent a broad range of age groups and social backgrounds. At the Museums Night 2011 an estimated 20,000 people visited a total of 176 events at 91 distinct locations, including exhibitions, galleries and interactive events. Events take place all over the city, mostly in the city centre, but some, such as the Museum of the MTU Aero Engines and the Potato Museum, are located in suburbs. The Science Night organisers reported [3] 28,000 visitors for 2011 with over 300 facilities offering in total about 1000 events. Both nights provide specially organised bus routes to transport visitors between events and in case of the Science Night also between the three cities (Nuremberg, Fürth and Erlangen).

These are large events where the user has huge choice of things they can do and limited time to see these things in. This is a prototypical casual-leisure need, whereby the goal is to maximise the pleasure of the user on the given evening. In such contexts information systems should try to lessen the burden on the user, assist in finding events that they will enjoy visiting and help construct a suitable itinerary for the chosen events.

4. LEARNING ABOUT VISITORS EXPECTATIONS & NEEDS

As a first step towards learning how to best support visitors in these contexts, we wanted to gain insight into what people want or expect from the Long Nights. To this end we conducted interviews with visitors at a central location on both nights (LNMuseum: n=25, LNScience: n=22). Reflecting event visitors generally, the interviewees were a diverse group, spanning different ages, genders and social backgrounds. The interviews lasted between 5 and 10 minutes and were recorded if interviewees gave permission, otherwise notes were taken. Questions revolved around what people wanted from the evening, the events or type of events they hoped to visit, how much they knew about events on offer and how they discovered / planned to discover events to visit.

From these interviews we know that on average each visitor plans to attend 4 events (median: 4) on the LNMuseum and 12 events (median 9) on the LNScience, which if true, will mean that approximately 80,000 visits (LNScience: 336,000 visits) will have taken place in 2011.

The interview recordings were transcribed and analysed qualitatively using an affinity diagramming technique, a group-based process, which allows the discovery and validation of patterns in the data [13]. This process consists of two stages. First, a brainstorming session is conducted whereby group members explain the observations they make in the data. We highlighted snippets from interview transcripts that we felt were informative with respect to our research goals and printed each separate snippet on a small piece of paper and used a large meeting table to identify observable patterns. The second stage involves finding a structure in the data by categorising and naming the responses. The process was conducted in a bottom-up fashion, with duplicate, similar or related responses being grouped together and the groups collapsed until a hierarchical structure was formed. It is important to note that we did not begin the process with a pre-defined model but allowed a coding scheme to emerge organically and inductively from the process. The approach taken aligns with the guidelines from grounded theory [12].

To test the coherency of the taxonomies, 3 coders (1 of whom did not participate in the categorisation creation process) re-coded 50 randomly selected snippets from the dataset. The assessors achieved substantial agreement (attaining a Fleiss Kappa score of 0.84[4]) on the top level of the hierarchy and high agreement on the top and second level(Fleiss Kappa = 0.72) .

Due to space constraints we cannot print the whole coding scheme hierarchy but in Table 1 we present the top level categories along with examples. These top-level needs fit well with the high-level, hedonistic needs reported in the literature [9, 31]. For example, typical desires were to have fun or enjoy an entertaining evening, to escape monotonous everyday life, to meet and spend time with friends or to re-live pleasant experiences from the past.

Category	Example
Wide range of events	"Events covering all different topics"
Social	Families adapting to their children
Escape everyday life	"Escaping from the everyday TV evening"
Culture	"Experience culture firsthand"
Learning	"Get known to new museums"
Describing experience	"Having fun on the evening"
Past experience	"Remembering old university days"
Event properties	"Not too far away"
Concrete event	"Want to visit Deutsches Museum"
Novelty	"See sth. that normally isn't available"

Table 1: Top-level of the coding scheme for visitors expectations

It is difficult to support many of these needs directly in a system. These need to be supported indirectly by assisting the user to find events that they will find fun, entertaining, informative etc. We focus on the properties of events that people suggested might help achieve these positive outcomes. Table 2 presents the "Event Properties" category in more detail, showing the 5 main kinds of attributes that participants reported as being important in the interviews.

Category	Example
Place/Travel time	"Don't want to spend too much time on the go"
Event type	"Don't like events with a fixed schedule"
Event topic	"I'm mainly interested in chemistry"
Take-part events	"Want to have an event where I can take part in"
Waiting time	"Don't like events with only 30 visitors at a time"

Table 2: Codes for category "Event properties"

From the interviews we also discovered that some visitors had made some efforts to plan their evening before the event while others made decisions regarding what to visit spontaneously on the night. Visitors typically discover events on offer using a booklet that is distributed for free by the organisers containing descriptions of all events in the order they lie along the bus routes. This booklet is necessarily large (LNMuseum: 110 A6 pages, LNScience: 256 A6 pages) and can be difficult to navigate.

[3]http://www.nacht-der-wissenschaften.de/2011/

[4]Fleiss' kappa is a statistical measure for assessing the reliability of agreement between a fixed number of raters when assigning categorical ratings to a number of items

The interview data further revealed that different people place different weight on event properties and any system should be flexible to support the varying needs of different users.

5. SYSTEM

For these Long Nights we developed two near-identical Android apps to help visitors find events of interest to them personally, accounting for the event properties listed in Table 2. Once they have found and indicated the events they would most like to visit, the system can create an itinerary for the evening, taking into account constraints such as start and end times of events and travel time between events based on public transport routes and schedules. If the user chooses more events than would fit into the available time[5], then the system tries to maximise the number of scheduled events by leaving out those that require long travel time. It is also possible for the user to manually customise plans by adding, removing and re-ordering events. Based on the created plan, the application can guide the user between chosen events using a map display and textual instructions. Figure 1 provides some screenshots of the app[6].

The user has four ways to find events he would like to visit by choosing one of the 4 tabs depicted in Figure 1:

The first tab (*by Tour*) allows the user to browse events organised geographically by the their position on bus routes laid on by the organisers. This is the same order as events are printed in the free booklet provided. This tab supports users in finding events based on location and may provide one way of minimising travel time between events - two attributes that were often mentioned as being important in the interviews.

The second tab (*by Genre*) provides a means to browse events by type or by discipline. The organisers provided different event metadata for each Long Night: event type e.g. "exhibition" or "guided tour" for the LNMuseum and disciplines for the LNScience. Ideally we would have provided both options on both nights, but lack of data made this impossible.

The third tab (*by Search*) provides free-text search functionality over the names and descriptions of the events. In a first step the event descriptions and titles were tokenised and stemmed. To match topically similar words we then map every token to one or more topic groups (these groups are taken from [7]). This way terms such as "dinner" and "food" are mapped to the same groups, thus event descriptions containing one of these words could be found by the other. The search tab, therefore, offered support to find events on topic, type of event, event descriptors or the title of the event itself. To speed up interaction with the system, queries were submitted after each typed character (search-as-you-type). More detailed descriptions of the search algorithms used can be found in [22].

The final tab (*Recommender*) offers personalised event recommendations to the user, generated from a hybrid recommender that combines content-based and collaborative filtering algorithms. Content-based algorithms require a user-profile, which the user could specify in our app via a slider interface to define their interest in different event types (LNMuseum) or disciplines (LNScience). A score is calculated for each event given the user profile using a simple scalar-product approach. The collaborative filtering algorithm mimics the natural human behaviour of making decisions based on recommendations from peers by utilising events selected by the user and also the event selections of users of the system. The ratings of all users are transferred to a central server that calculates a user-specific model as reported in [16] and sends it back to the client. On the client side, a score is calculated for each event based on the model. The score of the content-based recommender and of the collaborative filtering are then combined by weighting both scores with a confidence value. The confidence of the content-based one is determined by looking how diverse the different dimensions of the profile vector are. The confidence of the collaborative filtering approach is based on how many events the user has already selected.

The presented event lists in each tab contain the name and nearest bus stop for each entry. Users could tap events for further details and rate events for possible tour inclusion. The same interface is used when users want to add events to an existing tour or wish to go directly to one event.

In the remainder of the paper, in line with the research aims as outlined above, we focus on the way users found events of interest to them and properties these events had. All events can be found via any tab. Thus, if we can establish that events chosen from different tabs had varying characteristics, it may tell us something about what the users want when using this tab and help us judge the utility that these tabs offer.

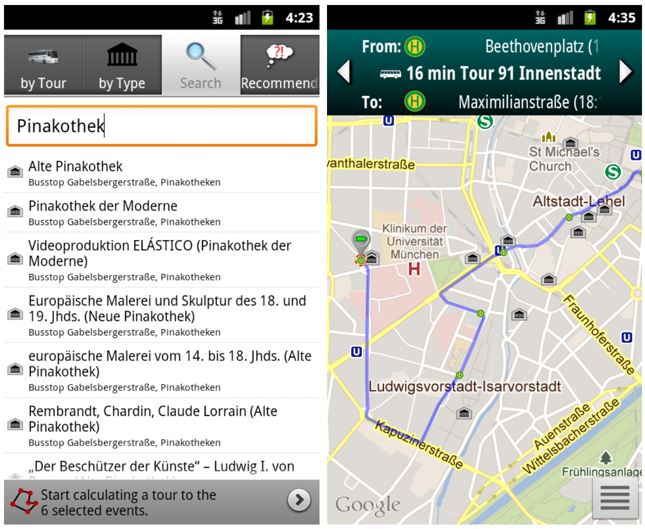

Figure 1: The search screen with a query (left) and the map screen with the planned route (right)

6. DATA COLLECTION

We examined user behaviour by recording user interactions with our apps at both Long Nights in 2011. The apps were on offer for download from the Android Market and advertised on the official Long Night of Museums and Long Night of Science web page. In total the LNMuseum application was downloaded approximately 500 times and 391 users allowed us to record their interaction data.

[5]most events are open between 7pm and 2am

[6]a video demo of the application can be found on YouTube (http://www.youtube.com/watch?v=woVjpivxtMc)

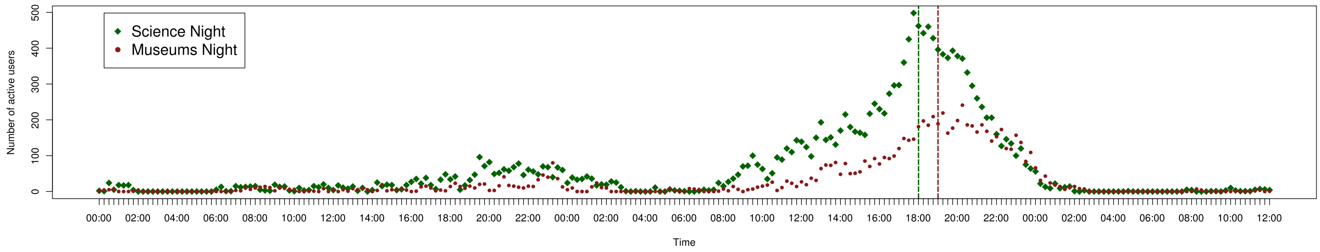

Figure 2: Number of active users within a 15 min timeslot on the day before the night till noon on the day after the night. Vertical lines mark the start of the night at 6 pm and 7 pm resp.

For the Long Night of Science our app was downloaded approximately 1000 times and 740 users gave permission for logging. We recorded all interactions with the application including tab changes, tour/event type/discipline selection, click-throughs, tours generated, modifications to tours, as well as all ratings submitted for events.

A short questionnaire provided us with demographic information. 53% of the app users were first-time visitors to the Long Nights, 21% were second time visitors and 26% had attended more than twice previously. 5% of users were 17 years of age or younger, 46% were between 18 and 29, 26% 30-39, 15% 40-49, 6% 50-59 and 1% above 60 years old. These demographics are very similar to those reported by event organisers for a previous Long Night of Music(LNMusic) in Munich [1] and suggest that our sample of users should reflect well the visitors as a whole. Comparing the age distributions for the Long Nights separately with Fisher's exact test against the demographics from the organisers reveals a p-value of 0.29 and 0.02 respectively. This shows that the demographics for LNMuseum app users is likely drawn from the same distribution as to the general visitor demographics reported for the LNMusic, however the users at the LNScience night follow a different age distribution. This likely reflects the much larger proportion of students who attended the science night and also, to a certain extent, the difference in regional demographics between Munich and the area around Nuremberg.

7. GENERAL USAGE OF THE SYSTEM

The application was used fairly intensively for both Long Nights. Users interacted with the system for 16.52 minutes (median 10.61) during the LNMuseum and 19.09 minutes (median 11.51) during the LNScience[7]. 67.26% of the LNMuseum users and 69.59% of LNScience users interacted with the system for more than 5 minutes; 15.35% and 20.81% respectively for more than 30 minutes.

Figure 2 shows the intensity of system usage before during and after the Long Nights. The graph generally conforms to what one might expect: usage increases rapidly from the morning before the Long Nights and peaks around the start times, finally tapering off again as events close. There is also a small ridge of activity on the day before the Long Nights, which peaks around evening time, indicating where people are perhaps planning their itineraries for the following day.

Before investigating usage of individual tabs we first obtain an overview of general usage. We do this by examining the distribution of "successful operations" across tabs. We define a *successful operation* as being either a rating of an event as a candidate for tour generation, a selection of an event the user wishes to visit immediately or insertion of an event into a pre-existing tour. Viewing of event details (click-through) is not included in this definition since this would be biased towards the two browsing tabs where viewing many events can be considered normal usage. In total 2,642 successful operations were performed on the LNMuseum and 4,646 on the LNScience. Table 3 shows the break down of these across the four tabs.

	by Tour	by Genre	Search	Recommender
LNMuseum	67.2%	12.1%	17.9%	2.8%
LNScience	53.9%	20.0%	23.0%	3.1%

Table 3: The % of succ. operations from each tab

The figures show that the *by Tour* tab was used much more heavily than the others. This is likely caused by a number of factors. Firstly, the *by Tour* tab is the default tab shown when first using the system. Secondly, the order of events in this tab corresponds to the order in the paper booklet. Thus visitors who marked interesting events in the booklet who may wish to "transfer" these into the app will most likely use this tab. Finally - based on the interviews - this tab can be expected to be the most frequently used since the location of an event was identified as one of the most important properties.

There is a big difference between usage of the *by Genre* tab for the LNMuseum and for the LNScience. As explained in Section 5 on the LNMuseum only the event type is available whereas on the LNScience the specific scientific discipline of the event is used. Although mentioned in the interviews, the event types defined for the LNMuseum are perhaps not so useful to users as they do not describe the content of the events but rather how it is organised or run. In addition on the LNScience there were a lot more individual events and bus routes, thus making a break down by discipline a much more attractive option.

Surprisingly few events were selected via the recommendations tab. This is especially surprising given that we ob-

[7]These figures were calculated by summing the time periods for which a user was active, discounting times where the system reported no interactions for more than 15 seconds. We further discounted any interaction sequence that contains gaps of non-interaction longer than 30 minutes as these are likely due to logging problems caused by running out of power, connection problems, app crashes, etc.

served very good results when we measured its performance with ratings data from a previous Long Night and found very accurate predictions.

We know from this overview analysis that all of the tabs were utilised at some point to select events. However we wanted to establish if it was the case that individual users each used a single specific tab from which to select their events, or if users made use of multiple tabs to build their itineraries.

Table 4 shows that about half of the users only added events via one tab while the other half of users used two or more tabs. If it is true that users made use of specific tabs to find events with a particular property, this suggests that around half of the users have more than one event property of importance to them.

	1 tab	2 tabs	3 tabs	4 tabs	1+2	1+2+3
LNMuseum	46.8%	39.0%	11.9%	2.2%	85.9%	97.8%
LNScience	56.6%	29.6%	12.6%	1.2%	86.2%	98.8%

Table 4: Number of tabs used by an user

We wanted to establish whether, in the event that multiple tabs were used, if one tab dominates and if so, by how much? To investigate this we calculated on a per-user basis the most frequently used tab, the second most frequently used and so on. The results of this analysis are presented in Table 5 and again indicate that people do not often switch tabs. The majority of events are selected from only one tab, seldom from a second tab and even more seldom from the third or forth tab.

	LNMuseum	LNScience
from most popular tab	84.3%	87.6%
from second most popular tab	13.2%	10.7%
from third most popular tab	2.2%	1.6%
from forth most popular tab	0.3%	0.1%

Table 5: Percentage of events from popular tabs

Table 6 shows the break down of tab popularity based on how many users had that tab as his favourite. These numbers are similar to the number of successful operation per tab (see Table 3). While there is a good spread across all tabs, the *by Tour* tab is clearly dominant.

	by Tour	by Genre	Search	Recommender
LNMuseum	60.6%	13.4%	23.8%	2.2%
LNScience	48.7%	20.7%	26.6%	4.0%

Table 6: Users with dominant tab

In this section we have analysed log data to show that the system was actually used, that the different tabs were used with varying frequency and that different users had different preferred tabs. All of this points to varying patterns of usage across the user population. In the following section we build on this to look at how the users interacted with the different system features (tabs) to find events.

8. USAGE OF SYSTEM FEATURES

8.1 Search system

Figure 1:left shows a screenshot of the search tab containing an input text box for query input. To ease input and save on precious space afforded by the small phone screens a search button was omitted and a search-as-you-type method used. To allow analysis it is therefore necessary to preprocess the recorded queries to establish those that the users actually intended to submit. For example, if the user wanted to search for "food", the system logged "f", "fo", "foo", as well as "food". Furthermore, should the user wish to submit a new query, then he must first remove the old search terms from the search box, resulting again in all prefixes but this time in decreasing length.

Automatically extracting the intended query proved difficult due to spelling errors and automatic correction. We therefore manually judged queries to be intended or not. Three assessors separately annotated all of the 10,000 queries logged on the LNMuseum and 21,000 queries logged on the LNScience as being either intended or not-intended. A high inter-assessor agreement was found (Fleiss' kappa = 0.872 and 0.899, 86.2% and 87.6% of queries which were labeled by at least 1 assessor were also labelled by at least one other assessor). This process resulted in a final list of 801 and 1,638 search queries, which is used in the following analyses.

8.1.1 Query Characteristics

Overall the search queries were short, having a mean length of only 1.21 terms ($\sigma = 0.52$) and 8.9 characters ($\sigma = 5.31$) on the LNMuseum and 1.26 terms ($\sigma = 0.71$) and 8.9 characters ($\sigma = 6.09$) on the LNScience. These values are much shorter than those reported for similar mobile-like devices for web search. [17] report lengths of 2.3 terms for older mobile phones and new research suggests even longer queries (2.9 terms and 18.25 characters) for modern phones similar to those used in our study [18].

It was very apparent while analysing the queries that many represented searches for named entities, in particular the names of specific museums. Again 3 human assessors were asked to assign queries into categories: specific event name, not a specific event name or indeterminate. The third category was necessary as some queries were short and it was not possible to definitively claim that the term referred to a specific event. For example "deutsches" is likely to be a reference to "deutsches Museum" but it is not possible to say for certain. The test was only done for the LNMuseum but queries from the LNScience look similar. For 87.3% of all queries at least two of the assessors were able to agree on one of the three categories (Fleiss Kappa of 0.43). 59.4% of the agreed on queries were marked as clearly named entities and 34.6% that might be named entities. Only 6.0% were labeled as non named entity searches. These remaining searches were often queries for non-museum locations, e.g. 18.2% of these are names of bus stops. Notably absent from the logs were queries describing topical content of events e.g. "art history", "engineering", "modern art", etc. There were also no queries referring to properties of events e.g. "interactive", "talks", "discussions" and no evidence of high-level, hedonistic qualities an event might bring about e.g. "fun", "exciting", "entertainment", etc. In line with previous query analysis papers, we analysed the diversity of submitted queries. The cleaned query set contained 417 unique queries on the LNMuseum and 943 unique queries on the LNScience. As expected the distribution looks rather Zipf-like with the top 2 queries being "deutsches" and "deutsches Museum" on the LNMuseum and "dolby" and "pathologie" on the LNScience. The top 50 unique queries amount to

43.1% and 28.3% resp. of all queries, the top 10 amount to 16.6% and 11.9% resp. and the most common search term was used in 2.5% and 1.6% resp. of all searches.

A notable observation is that the queries submitted to the search system do not seem to reflect the information needs described in the pre-study interviews. It seems as if the users did not use the search engine to discover new events that matched desired properties (i.e. those in Table 2), but rather used the feature to filter to events they already knew existed. Reflecting this, our queries have similar properties to those reported for known-item searches in web, email and desktop search, which have also been shown to be very short and contain a high percentage of named-entities [8, 27].

8.1.2 Query Performance

To understand how successful queries were we looked at the users' interactions with search results. A metric equivalent to click-through data refers to whether the user selected a returned result to read a detailed description of the event. On the LNMuseum 58.4% of all searches resulted in a click-through with an average of 0.73 clicks per query ($\sigma = 0.93$) and 5.95 results on average ($\sigma = 9.10$). Similar on the LNScience 63.2% of all searches gave rise to a click-through with an average of 0.86 clicks per query ($\sigma = 1.63$) and 9.15 results on average. We didn't consider good abandonment since the result list contains no information beyond name and nearest bus stop.

When considering our successful operations metric defined in Section 7, we found 53.6% of the LNMuseum queries and 55.4% of the LNScience queries to be successful. A large percentage of queries (40.3% and 34.0% respectively) were unsuccessful, and 59.8% and 53.9% of those were using a search term which resulted in an empty result list. This was in most cases a miss-spelled or only partially-written named entity. The huge number of spelling errors underlines the potential benefits of using fuzzy search methods in this application context.

8.2 Browsing behaviour

There are two tabs where users can browse through the events: The *by Tour* tab, where events can be filtered by transportation link, and the *by Genre tab*, where events can be filtered by event type or discipline. We want to analyse whether users used these filtering mechanisms to efficiently cope with the information overload caused by the vast amount of events on offer to them. We, therefore, looked at the number of filters applied on a per user basis. Table 7 shows the percentage of users that used one, two, three, four or more filters. Again usage is defined as at least one successful operation on that tab and only users with at least one are considered.

used tab	avail. filter	number of used filters 1	2	3	4	>4
by Tour	6	23.9%	25.8%	17.8%	14.7%	17.8%
by Genre	14	36.1%	37.7%	19.7%	4.9%	1.6%
by Tour	13	41.0%	28.1%	11.9%	7.6%	11.4%
by Genre	31	29.0%	26.0%	14.0%	17.0%	14.0%

Table 7: Used filters on browsing tabs (top rows: LNMuseum; bottom rows: LNScience)

Comparing the *by Tour* tabs it is notable that even though on the LNMuseum fewer bus routes were offered than on the LNScience, a far higher percentage of users tended to combine multiple filters for the LNMuseum. Looking at the *by Genre* tab we see the expected result; more available filters lead to more diverse usage thereof. Interestingly users used only a very tiny fraction of the available filters. Thus they tended not to browse through all the events, but rather narrowed the amount of events down to a manageable size using attributes of interest to them.

The analyses performed so far suggest that the different tabs were used with very different purposes in mind. The *Search* tab, contrary to its design, was used mainly as a lookup tool for events that were already known about; the browsing tabs, i.e. *By Tour* and *By Genre* were the main method of event discovery, with users narrowing down the search space by some criteria important to them; and the *Recommender* tab, which was specifically designed to facilitate the discovery of events tailored to user preferences was rarely used.

In the following section we explore this hypothesis in more detail by looking at the properties of events that were selected. We wish to establish if the various tabs on offer did indeed lead to different events being chosen and if those events had properties reflecting the design aims for each tab.

8.3 Event Properties

We examine the events chosen in terms of 4 metrics related to the aspects reported as important in the interviews (see Tables 1 and 2). For each tab, we look at how dispersed the events were both spatially and temporally, which accounts for the place / travel code in Table 2. We examine how topically diverse events were as some interviewees reported that diversity was important. Finally, we analyze the popularity of selected events with respect to how often events were selected over all tabs. This is related to event novelty.

8.3.1 Spatial contiguity of events

With the *By Tour* tab designed to assist people in selecting events by location we want to see what influence the usage of this tab has on the length of the journey they take. We therefore define the *spatial contiguity* of a set of events as the length of the shortest path connecting all events as the crow flies. This problem is similar to Traveling Salesman Problem but instead of a round trip an open path is calculated. To ease calculation we use a very simplistic greedy approximation algorithm: A path is gradually extended by adding the event to the path with the smallest distance to one of the ends of the pre-existing path. To accommodate for different set sizes we normalise the travel distance by the number of connection parts (number of events − 1).

We compare two (not necessary distinct) sets of users. The sets contain all users that performed at least two successful operations on either the by Tour tab (first set) or any of the other tabs (second set) and where the spatial contiguity of these events was greater than 0. In Table 8 the number of users and the average normalised travel time per set is shown.

The normalised travel distance is about the same for both sets on the LNMuseum. However, for the LNScience the normalised travel distance is about half as long when events were chosen from the *by Tour* tab than when they were chosen from the other tabs; a highly significant difference. We expect this to be the case because on the LNScience three

	LNMuseum	LNScience
Users using by Tour tab	158	198
Normalised travel distance	$\mu = 1431.95m$	$\mu = 1382.86m$
of events from by Tour tab	$\sigma = 1752.47$	$\sigma = 1925.94$
Users using other tabs	136	216
Normalised travel distance	$\mu = 1499.18m$	$\mu = 2646.22m$
of events from other tabs	$\sigma = 1403.79$	$\sigma = 2865.78$
difference of travel distance	$67.23m$	$1263.36m$
p (Gauss test)	0.715	$1.139009 \cdot 10^{-7}$

Table 8: Spatial contiguity of events

cities are involved covering large areas without any events in between, whereas on the LNMuseum events are concentrated on one city. Thus the distance between events chosen from different cities leads to larger travel distances.

The large difference observed between the *by Tour* tab and other tabs is perhaps not surprising given that the bus routes are designed so that together they maximise coverage over the area of events. This means that if a user only selects events covered by a single route then he will only be visiting events within the small portion of the total area covered by all the routes. If however he selects events covered by multiple bus routes then it is more likely that they will be more widely spread apart.

8.3.2 Temporal contiguity of events

From the interviews on the LNScience we know that a lot of people consider travel time to be much more important than travel distance. Thus we performed another test regarding contiguity but this time considering travel time. We define *temporal contiguity* of a set of events as the time needed for a route utilising all available transportation links and visiting all events of that set. This is similar to the *spatial contiguity* but using a different metric for distances. This metric is normalised by dividing through by the number of connecting parts of that route (number of events -1).

We again want to exclude users that visit only one location and thus included only users with at least two successful operations from either the *by Tour* tab or the other tabs and that have an temporal contiguity of at least one minute. Table 9 shows the normalised traveling time from both sets.

	LNMuseum	LNScience
Users using by Tour tab	157	198
Normalised travel time	$\mu = 14.34min$	$\mu = 14.41min$
of events from by Tour tab	$\sigma = 10.80$	$\sigma = 14.32$
Users using other tabs	133	210
Normalised travel time	$\mu = 19.25min$	$\mu = 27.54min$
of events from other tabs	$\sigma = 11.92$	$\sigma = 23.65$
difference of travel time	$4.91min$	$13.12min$
p (Gauss test)	0.000262	$8.88178 \cdot 10^{-12}$

Table 9: Temporal contiguity of events

Events chosen from the *by Tour* tab are significantly more temporally contiguous than the those chosen from the other tabs. For the LNMuseum people save around 5 minutes per connection, which considering visitors chose 4 events (3 connections) on average for the whole night (see Section 8.3.3), sums up to a total saving of 15 minutes. For the LNScience the time saving is even greater at 13 minutes per connection - which we again attribute to the three cities involved - leading to a expected total saving of about $1\frac{3}{4}$ hours, considering the reported 9 visits per evening.

In summary the first tab can be considered as an effective assistance feature that helps people to find contiguous events (both spatially and temporally), allowing them to spend more time at events and less time traveling between them.

8.3.3 Diversity of events

The *by Genre* tab allows users to filter events by different genres (LNMuseum: by event type; LNScience: by discipline). As interviewees reported very different needs regarding type or discipline - some preferred to have a wide range of events, other wanted a very focused selection - we were interested how the diversity of the genres of events chosen from this tab compares to ones chosen from other tabs. *Diversity* of a set can be measured in a number of ways, we opted for two commonly-used metrics: First we report the *Shannon entropy* [24] of the users' choice for events of specific genre[8]. It is calculated by

$$H = -\sum_{i=1}^{G} p_i \log p_i \qquad \text{with } p_i = \frac{n_i}{n}$$

where G is the number of genres, n_i denotes the number of chosen events from genre i and n is the number of chosen events in total. H is a measure of the mean self-information of the chosen genres. In our context higher values mean a higher diversity of genres. As the entropy is difficult to judge by humans and unbounded we also report the *Simpson index*[25] as another diversity metric which is defined as

$$D = 1 - \sum_{i=1}^{G} \frac{n_i \cdot (n_i - 1)}{n \cdot (n-1)} \quad .$$

D is an estimator of the probability that two drawn events from the set belong to different genres. It is 0 for no diversity and approaches 1 as the diversity of the set increases towards complete diversity. We compare two sets containing all users that performed at least two successful operations on either the *by Genre* tab (first set) or any of the other tabs (second set). In Table 10 both diversity metrics are shown for the two sets for both nights.

	LNMuseum	LNScience
Users using by Genre tab	61	100
Average Entropy	0.848 bit	1.375 bit
Simpson Index of event	$\mu = 0.61$	$\mu = 0.72$
from by Genre tab	$\sigma = 0.34$	$\sigma = 0.29$
Users using other tabs	221	323
Average Entropy	1.298 bit	1.674 bit
Simpson Index of event	$\mu = 0.77$	$\mu = 0.83$
from other tabs	$\sigma = 0.21$	$\sigma = 0.21$
Wilcoxon Rank-Sum-Test		
on Simpson Indices	0.0002128	$3.795 \cdot 10^{-5}$

Table 10: Diversity of events

The results show that both diversity metrics are smaller for the *by Genre tab.* A Wilcoxon signed-rank test performed on the Simpson index results proved this difference to be highly significant, indicating that users of this tab have very specific interests they want to focus on. In contrast, it seems that users that wish to see a wide range of different events generally avoid this tab. When comparing filtering by event type (the LNMuseum) and by discipline (the LNScience) an 11% higher diversity can be found when filtering by discipline. This difference may stem from there being more than twice as many filtering options of disciplines (31) for the LNScience compared to only 14 event types for the LNMuseum.

[8]In our system events could belong to multiple genres

A test with both filtering options available on the same night is needed to further clarify the verity of this outcome.

8.3.4 Popularity of events

In the interviews visitors often reported that they wished to see specific events, with many different visitors mentioning mostly the same small set of very popular events. On the other hand, interviewees also wished to avoid events they expected to be crowed and some wanted to visit novel, previously unknown events. We do not have a direct measure for these properties, however they can be seen as being related to popularity. Novelty can be viewed as the opposite of popularity. We therefore analyse which tabs were used for popular events and which were used for less popular events. Especially interesting is the recommender tab, even though it was the least often used. Collaborative filtering is reported to tend to popular items [26] whereas there is no reason why a content-based recommender would have such a bias.

To obtain a measure indicating the relative popularity of an event we count the number of successful operations conducted on that event. This yields a Zipf-like distribution with the top 10 events accounting for 21.23% of all successful operations for the LNMuseum and 10.85% for the LNScience. To calculate the share of popular events from a given tab we normalised the choices for each individual event by the total number of successful operations performed on that tab. We then accumulate these numbers for the (overall) Top N events which is shown for the LNScience in Table 11. As can be seen for the LNScience the Top 50 events account for 41.12% of successful operations on the *Search* tab. Following *Search* are the recommender tab, the *by Tour* tab and finally the *by Genre* tab. This ranking of tabs also holds true for other values of N except for very small N. In general, popular events were more likely to be selected from the search and recommender tabs than from the browsing tabs. Results from the LNMuseum - shown in Table 12 - are closer to each other and very dependent on N, most likely due to the smaller number of events and users. Interestingly, the search tab is used considerably more for the Top 5 and Top 10 popular events but for the popular events outside of the Top 10 the recommender tab is similar in usage. From the analysis of search queries we know that people mainly performed known-item search which is only possible for events people know of in advance. We believe this generally only holds true for the most popular events. Further studies would be necessary to obtain better data, allowing a more detailed comparison between the recommender and search tabs. The usage of both browsing tabs seem to be independent of the event popularity as they are similar to the overall popularity of all tabs combined. This result is not surprising and is likely due to the fact that both tabs contain static lists of events.

	all	by Tour	by Genre	Search	Rec.
Top 5	6.37%	6.11%	3.66%	8.97%	**9.15%**
Top 10	10.85%	10.82%	7.64%	**13.55%**	11.97%
Top 20	18.12%	17.52%	13.56%	**23.55%**	17.61%
Top 50	34.40%	33.37%	28.85%	**41.12%**	38.03%

Table 11: Popular events on the LNScience

	all	by Tour	by Genre	Search	Rec.
Top 5	12.76%	12.62%	12.85%	**13.53%**	10.67%
Top 10	21.23%	20.68%	22.26%	**23.47%**	16.00%
Top 20	34.25%	34,08%	33.54%	34.25%	**41.33%**
Top 50	62.30%	61.41%	61.44%	**65.75%**	65.33%

Table 12: Popular events on the LNMuseum

9. DISCUSSION AND CONCLUSIONS

The analyses above reveal several key findings:

The system was used by participants. The developed app was downloaded and used by a large number of people (approx. 2% of the visitors to the Long Nights). The analyses revealed that the app was used throughout the evenings. From the logs we can infer that some users used it as a means to plan their evening in advance while others first made use of app when they were already in the city.

The *By Tour* Tab was most frequently used. All of the app tabs on offer were used by some users to add an event or rate an event positively. Users did, however, tend to select the majority of events from the *By Tour* tab. We offered several explanations for this including biases due to the position of the tab and the fact it was the default on starting the app. Analyses of the events selected from the tab suggest however other reasons. The geographical area covered by selected events was smaller when users chose from the *by Tour* tab, likewise travel time between the events was also significantly smaller. These two points were emphasised by interview participants as important when selecting events - users don't want to spend too much of their evening getting to and from events and therefore want to maximise their time actually attending.

Most users choose the majority of their events from only 1 tab and very seldom use multiple tabs. Our analyses reveal that users tended to switch between tabs very infrequently. Most users selected the majority of events from only one particular tab. However the choice of "favourite tab" differed among the users, although again, the *By Tour* tab was the clearly most frequently used. This points to there being different modalities of use, dependent on the user.

Different tabs seem to be used with different aims in mind. Both the way the various tabs were used and the properties of the events selected from them further hint that users had different goals in mind when using different tabs.

The *Search* tab was used almost exclusively for known-item search, with search queries being unusually short and mainly consisting of (partial) named-entities. Surprisingly given what people said during the interview, there was very little evidence found in the logs of people searching for general themes, topics or keywords.

The *By Genre* and *By Tour* tabs were the main source of event discovery. Nevertheless, there was a clear (and perhaps unsurprising) bias towards popular, well-known events in all tabs. Using the search features resulted in the most popular events being selected, again pointing to known-item search behaviour.

The *Recommendation* tab - the tab best placed to promote novel events that the user might like to visit - was used very infrequently. Although, for us, this was one of the most surprising findings, there could be several explanations. It could be, for example, that the users did not understand that the recommendations were personalised and instead treated this tab as some kind of top 10 list. The lack of usage may also have been a result of a trust issue; whereby the users did not believe that the recommendations provided could be useful. A third explanation could be that the limited usage of this feature was simply a result of a negative bias due to the recommender being the last tab. Due to these biases we believe that further studies would be necessary to ascertain how recommender systems can better be utilised for these events.

We believe the key takeaway from this research is that people have very different ways that they wish to interact with the data in this context. There is therefore a need for a variety of different methods for accessing, browsing and searching the data contained within the app. Out of the criteria revealed in the interviews, diversity seemed the most poorly catered for attribute in our app. Of particular note is that the main source of event discovery, the browsing tabs, provided the least diverse set of selected events.

10. FUTURE WORK

The investigations made in this paper point to a number of avenues for potential future work. One obvious next step is to analyse how usage of the system linked to a visitors experience of the evening. We plan to explore how different usages of the system related to metrics such as how many events were visited on the evening, how visited events were rated, how long visitors attended events and how they travelled. These are metrics that can be directly related to the user's experience of the event. Building on our finding that some people plan prior to attending events and some people make spontaneous decisions while attending, it would be interesting to see if this leads to different outcomes measured by such experiential metrics.

Further, as some of the system features were not used as intended (e.g. the search features) or used far less often than expected (e.g. the recommendations), in future Long Nights we plan to experiment with interface changes to see if the users can be encouraged to use the system as it was intended. For example, we want to determine if the user can be persuaded to submit topical searches rather than simply known items by having a greyed out example in the search box. Similarly, we will try out several versions of the interface to encourage use of the recommender system, for example, by changing this to be the default tab to counter the positional bias. A second idea is to try and build trust in the system by highlighting events in the search and browse interfaces that the recommender system predicts will be liked by the user.

The analyses presented in this paper inform on casual-leisure information seeking behaviour in a specific contextual situation. A third separate line of future work relates to learning about casual-leisure information seeking behaviour in other contexts to see if the trends uncovered in this work hold in different contexts i.e. do people have preferences for particular interaction modes when looking for music or television programmes and do these different modes lead to items being found with particular attributes?

11. REFERENCES

[1] *Die Lange Nacht der Musik Besucherbefragung 2010.* Münchner Kultur GmbH, 2010.

[2] N. J. Belkin, R. N. Oddy, and H. M. Brooks. ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2):61–71, 1982.

[3] K. Byström. *Task complexity, information types and information sources. Examination of relationships.* PhD thesis, University of Tampere, Dep. of Inf. Studies, 1999.

[4] K. Church, J. Neumann, M. Cherubini, and N. Olive. The "map trap"? an evaluation of map versus text-based interfaces for location-based mobile search services. In *WWW '10*, 2010.

[5] K. Church and B. Smyth. Understanding the intent behind mobile information needs. In *IUI '09*, 2009.

[6] B. Dervin and M. Frenette. Sense-Making Methodology: communicating communicatively with campaign audience. In *Sense-Making Methodology reader: selected writings of Brenda Dervin*, pages 233–249. Hampton, 2003.

[7] F. Dornseiff. *Der deutsche Wortschatz nach Sachgruppen.* DeGruyter, Berlin, New York, 2004.

[8] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff I've seen: a system for personal information retrieval and re-use. SIGIR '03, pages 72–79, NY, 2003. ACM.

[9] D. Elsweiler, S. Mandl, and B. Kirkegaard Lunn. Understanding casual-leisure information needs: a diary study in the context of television viewing. IIiX '10, pages 25–34, NY, 2010. ACM.

[10] D. Elsweiler, M. L. Wilson, and B. Kirkegaard Lunn. *New Directions in Information Behaviour*, chapter Understanding Casual-leisure Information Behaviour. Emerald Pub., 2011.

[11] D. Gavalas and M. Kenteris. A web-based pervasive recommendation system for mobile tourist guides. *Personal and Ubiquitous Computing*, 15:759–770, 2011.

[12] B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: strategies for qualitative research.* Aldine de Gruyter, New York, NY, 1967.

[13] J. Hackos and J. Redish. *User and Task Analysis for Interface Design.* John Wiley & Sons, Inc., 1998.

[14] K. Hagen, R. Kramer, M. Hermkes, B. Schumann, and P. Mueller. Semantic matching and heuristic search for a dynamic tour guide. In A. J. Frew, editor, *Information and Communication Technologies in Tourism 2005*, pages 149–159. Springer Vienna, 2005.

[15] P. Hansen. User interface design for IR interaction. a task-oriented approach. In *CoLIS 3*, pages 191–205, 1999.

[16] M. Harvey, M. J. Carman, I. Ruthven, and F. Crestani. Bayesian latent variable models for collaborative item rating prediction. In *Proc. 20th ACM CIKM*, CIKM '11, pages 699–708. ACM, 2011.

[17] M. Kamvar and S. Baluja. A large scale study of wireless search behavior: Google mobile search. In *CHI 2006*, 2006.

[18] M. Kamvar, M. Kellar, R. Patel, and Y. Xu. Computers and iphones and mobile phones, oh my!: a logs-based comparison of search users on different devices. WWW '09, pages 801–810, NY, 2009. ACM.

[19] A. Maruyama, N. Shibata, Y. Murata, K. Yasumoto, and M. Ito. P-tour: A personal navigation system for tourism. In *Proc. of 11th World Congress on ITS*, pages 18–21, 2004.

[20] A. M. Pejtersen. The book house: modelling users needs and search strategies as a basis for system design. Technical Report Riso-M-2794, Riso National Laboratory, 1989.

[21] C. S. Ross. Finding without seeking: The information encounter in the context of reading for pleasure. *IPM*, 35(6):783–799, 1999.

[22] R. Schaller, M. Harvey, and D. Elsweiler. Out and about on museums night: Investigating mobile search behaviour for leisure events. In *Proc. of Searching4Fun Wksp, ECIR*, 2012.

[23] M. V. Setten, S. Pokraev, J. Koolwaaij, and T. Instituut. Context-aware recommendations in the mobile tourist application compass. In *AH 2004, LNCS 3137*, pages 235–244, 2004.

[24] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comp. Com. Rev.*, 5(1):3–55, Jan. 2001.

[25] E. Simpson. Measurement of Diversity. *Nature*, 163:688, Apr. 1949.

[26] H. Steck. Item popularity and recommendation accuracy. In *RecSys '11*, pages 125–132, New York, NY, USA, 2011. ACM.

[27] J. Teevan, E. Adar, R. Jones, and M. A. S. Potts. Information re-retrieval: repeat queries in Yahoo's logs. SIGIR '07, pages 151–158, NY, 2007. ACM.

[28] J. Teevan, A. Karlson, S. Amini, A. J. B. Brush, and J. Krumm. Understanding the importance of location, time, and people in mobile local search behavior. In *Proc. of MobileHCI*, MobileHCI '11, pages 77–80. ACM, 2011.

[29] E. Toms. Understanding and facilitating the browsing of electronic text. *J. of Human-Comp. Studies*, 52:423–452, 2000.

[30] P. Vansteenwegen, W. Souffriau, G. V. Berghe, and D. V. Oudheusden. The city trip planner: An expert system for tourists. *Expert Syst. Appl.*, 38(6):6540–6546, June 2011.

[31] M. L. Wilson and D. Elsweiler. Casual-leisure Searching: the Exploratory Search scenarios that break our current models. HCIR '10, Aug 2010. New Brunswick, NJ.