A Field Trial to Elicit Individual Preferences in the Context of a Mobile Dynamic Tour Guide[1]

Ronny Kramer[a], Marko Modsching[a],
Klaus ten Hagen[a], and Ulrike Gretzel[b]
[a] Department of Computer Science
University of Applied Sciences Zittau/Görlitz, Germany
{K.tenHagen, MModsching, RKramer}@HS-ZiGr.de

[b] Laboratory for Intelligent Systems in Tourism
Texas A&M University, USA
ugretzel@tamu.edu

Abstract

Knowing tourists' individual preferences provides the possibility to offer personalised tours. The challenge is to capture these preferences using a mobile device. During a field study in Görlitz three methods for preference elicitation were evaluated. The results served to clarify fundamental questions en route to developing a personal tour guide: 1) Is it possible to seed interest profiles in the mobile context with all its distractions that allow the accurate prediction of rankings of concrete sights? and, 2) Are the interest profiles sufficiently diverse to justify the computation of personalized tours instead of selecting from a list of standard tours? The results suggest that choice among simple categories constitutes the most effective means of capturing user preferences and that standard tours would not adequately represent the wide array of interest profiles identified.

Keywords: dynamic tour guide, personalized tour, mobile computing,, semantic matching, elicitation of preferences, mobile recommender

1 Introduction

One goal of a Dynamic Tour Guide (DTG) – a mobile application enabling a personalized, spontaneous and guided tour – is to provide individualized tours and information to tourists by means of pervasive computing based on the actual context which is defined by personal interests, location, and schedule of a tourist (Ten Hagen et al., 2004; Abowd et al., 1997; Cheverst et al., 2000). The challenge lies in eliciting the preferences of a tourist by means of a mobile device to seed an interest profile, to

---

[1] Published in "Information and Communication Technologies in Tourism 2006", Andrew J. Frew, Martin Hitz, Peter O'Connor (Editors), Springer Computer Science

rank the available sights by these interests (semantic matching), to compute an individual tour based on these data, and to adapt the tour to spontaneous choices made by the tourist when executing the tour. The specification of interests in a mobile context is particularly difficult (Nguyen, Cavada & Ricci, 2004). A mobile device provides less than 4% of the pixels of a PC and its users are subject to many distractions, e.g. traffic noise, that make them less patient in interacting with the application. Thus, time and information bandwidth is severely limited compared to a standard PC environment. This paper suggests three different GUIs and preference elicitation models considering these constraints. The gathered interest profiles are subsequently benchmarked in order to determine the accuracy of the elicitation process. In addition, the diversity of the tourists' interests is analyzed in order to study the necessity to gather individual profiles.

2 Related work

A survey of visitors to the city of Heidelberg found that most tourists explore the city by foot and on their own (Freytag, 2003). Only 7% decide to experience the city through a guided tour. Further, almost all tourists visit the local castle while all other sights receive less attention; some are visited by less than 5% of the tourists. Kempermann et al (2003, 2004) report that spatial behaviour of tourists at theme park destinations can be highly concentrated; however, significant differences exist between first-time and repeat visitors. First-time visitors have less information about a destination and try to visit as many major attractions as possible in order to spread the risk of visiting an attraction that does not suit their preferences. In contrast, repeat visitors choose attractions more selectively because they already know what to expect. A DTG may help tourists visiting destinations for the first time in picking out the sights most interesting to them by providing important information that might otherwise not be accessible to tourists or at least not in a highly personalized form. In addition, personalized suggestions made by a DTG can expose repeat visitors to attractions not previously visited and might actually increase repeat visitation over time.

A precondition to making suitable recommendations is obtaining the necessary information about users' preferences. Gretzel and Fesenmaier (2005) suggest that the process of capturing user preferences is persuasive in and by itself. Most people are not aware of their preferences all the time, so they need to be inspired or reminded more so than exactly measured. Consistent with these findings, Gretzel et al. (2004) report that choice among simple categories can be an equally effective way of predicting preferences than lengthy questionnaires that aim at capturing detailed interest profiles. Nevertheless, eliciting user preferences is not a trivial task and there seems to be no general solution to the problem, at least not in a mobile context.

Solving the puzzle, however, is important as more personalized tour suggestions promise to increase tourists' ability to enjoy a destination to its full potential.

3 Methodology

A field study was designed to answer the following questions:

(1) Is it possible to build a mobile system that effectively collects general preference information from tourists and successfully recommends attractions?

(2) How diverse are the interests of tourists? Are the interest profiles so similar that the optimal tours hardly differ and offering a selection of standard tours would sufficiently serve the purpose?

The field study was conducted in Görlitz, Germany in the summer of 2005 over a time period of four weeks. Before the experiment could start, about 80 sights of the city of Görlitz were modeled semantically, which means that they were assigned to classes of an attraction ontology. In addition, pictures and describing text were collected for each attraction in order to provide appropriate means for the study subjects to rank the concrete sights.

For the experiment itself, subjects were approached in the city center of Görlitz and asked if they would like to help evaluate a new mobile technology. In the course of the study, 235 tourists were given an MDA, which is a PDA type handheld device with an integrated mobile phone. Their first task was to complete a self-administered questionnaire asking for age, gender and previous experience with certain computing and communication technologies. The study subjects were then randomly assigned to one of three preference elicitation conditions displayed on the MDA (Figure 1):

- **Hierarchical (Tree) Browser**: The hierarchical structure of the ontology is visualised by a tree view element. The user can select any category he/she is interested in by checking the corresponding boxes. The advantage is that everything can be displayed on a single screen which turns into a disadvantage at the same time as small fonts have to be used and scrolling becomes necessary when expanding the tree.
- **Main Categories**: Only the main categories are provided for selection. Selecting one category opens a pop-up window that invites users to provide a percentage value to express the intensity of their interest.
- **Inspirational Images**: The hierarchy is presented by single screens for each level. The concrete interest terms are displayed by text and additional exemplary images or icons. These images shall inspire associations causing positive or negative

feelings with each term as many people may not be able to imagine e.g. a baroque building by only reading it. But to provide fair and realistic selection options these images must be equally styled in order to be similar attractive themselves. The pictures can be maximized and detailed information for each term is offered. The advantage here is the visualisation by pictures and symbols. However these images make a lot of screens necessary and therewith lead to a difficult orientation between the levels.

Fig. 1: Preference Elicitation Methods

After the participants had expressed their interests using one of the three methods, the semantic matching algorithm rated all available sights appropriately. Two high rated, two medium rated and two less desirable sights were picked out by chance and displayed. The tourist was then asked to rank the six concrete attractions using descriptions and pictures provided for each. The purpose of this specific task was to measure the congruence between the ranking provided by the semantic matching algorithm and the ranking of the individual tourist. Interest profile data, rankings calculated by the system, as well as the rankings provided by each tourist were automatically captured and later analyzed using descriptive and multivariate statistics.

4 Results

The age of the 235 tourists ranged from 13 to 78, with an average age of 47 years and a modal age of 60 years. This age distribution corresponds to the overall visitor profile of Görlitz. Of those who participated in the study, 60% were male and more than two thirds stated they regularly used a PC. Still more than one third of the subjects reported they often use the Internet and mobile phones, while almost nobody uses an MDA. Almost all (90%) indicated they own a PC or a mobile phone.

4.1 Interaction durations

Table 1 shows the duration of the interest elicitation process using the MDA, the number of clicks, and the duration of a tourist's interaction with a single panel. Surprisingly for all three methods tourists spent about 2 minutes or less specifying their interests with about 9 seconds per screen using a total of 22 clicks.

Table 1: Preference elicitation: Duration, number of clicks and panel view

Method		**Tree**	**Images**	**Categories**
Duration of elicitation process [min]	Mean	2.03	2.12	2.03
	Median	1.44	1.28	1.50
Clicks	Mean	16	29	21
	Median	12	13.5	19
Duration of panel view [sec]	Median	5	9	8
	Min	2	3	4
	Max	10	20	13

The screens for the image version are the most complex ones providing pictures, text and scrolling opportunities; thus, they were viewed the longest. Yet, the median values and tendencies for all three methods hardly differ and potentially define the attention span of a tourist of this age in a mobile context.

The distribution of clicks used by tourists to specify their interests shows that the tree and category method have one well defined peak, whereas the image method shows two of them. Many tourists remained in the top-level, indicated by the first maximum, and therefore needed very few clicks. But those who also used the deeper levels needed the most clicks of all, which is shown by the second maximum. Not surprisingly the number of clicks and the duration of the interest gathering process are closely correlated with 0.82. The average time between two subsequent clicks ranges from 7.6 seconds for the tree view and 4.4 seconds for the image version. The number for the image version is the smallest, since many clicks were executed for scrolling to view the entire choice.

4.2 Interest selection

Across all three preference elicitation conditions, the category "architecture" or its subcategories were chosen most often. This result was expected since Görlitz offers many architectural sights that attract the majority of the tourists. Landmarks were the second most frequently selected category.

The most important question in this context is how detailed the interests of tourists are and how detailed they are willing to specify them. The system offered five main categories. Close to 40% of the tourists selected interest terms out of at least 2 different categories and about 4 different ones in general. The diagram in Figure 2 displays the deepest levels the tourists reached when specifying their interests for the tree and image version. For both the image and the tree method, a selection of interests at deeper levels was possible. And indeed, more than 70% of the tree explorer users and 50% of the image version users made use of this option and specified more detailed interests beyond the top-level categories.

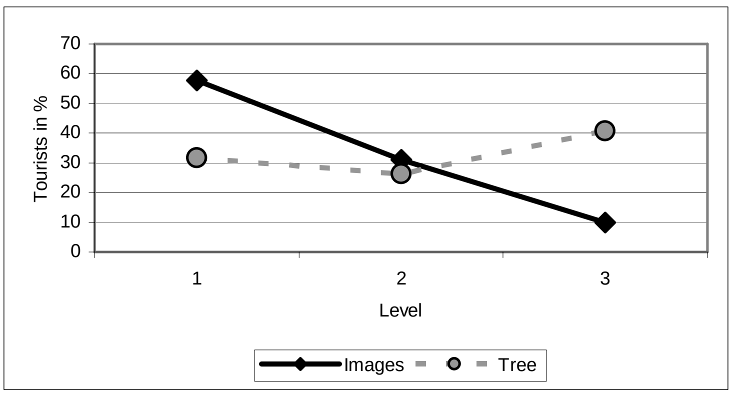

Fig. 2. Distribution of interest specifications.

But this analysis also shows that almost half of all tourists using the inspirational image method did not specify interests at any level deeper than the first. The usage pattern was thus very similar to the much simpler method using main categories as they mainly selected top-level categories.

The data further indicates that most tourists did not realize that clicking on a category name or its icon reveals subcategories displayed by concrete images, though it was mentioned on the screen. The problem of the inspirational method on a mobile device is that in order to provide choice at least two images need to be shown. To minimize scrolling, four images were selected; however, using four images does not leave space to include visual clues to the hierarchical nature of the selection.

4.3 Rank Order Correlations

The similarity between user defined rankings and algorithm-based rankings can be expressed in the form of rank order correlations. The best result is an identical ranking (value 1), the worst one is an opposite list (value -1). A value of 0 signifies that there is no recognizable correlation. The correlation results are listed in Table 2. The results illustrate that the relatively simple method using five main categories and the inspirational method using images are equally effective in capturing the interests of the tourists. The least effective method is the Windows Explorer style hierarchy browser. Figure 4 shows, that more than half of all tourists have reached a correlation higher than 0.5, which means that for the majority of the tourists the recommendations are pretty good. Only very few reached a negative correlation (Figure 3).

Surprisingly the amount of clicks as well as the amount of time does not have a positive effect on the Spearman coefficient. The correlation for both dependencies is -0.1, which means that there is little dependency. One might now reason that 2 minutes must be enough to elicit preferences from a tourist. At least with the three methods used in this field study any further effort does not lead to an improvement of the ability to predict the selection of attractions.

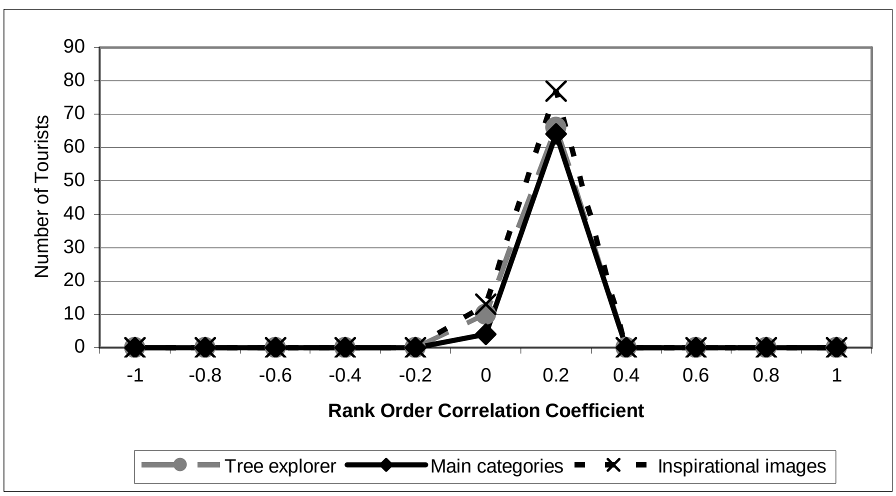

Fig. 3. Preference elicitation: Correlation

Table 2: Preference elicitation: Correlation

Method		Tree	Images	Categories
r_s coefficient	Mean	0.47	0.48	0.52
	Median	0.54	0.60	0.60

4.4 Entropy calculation

Given the values of the rank order correlation coefficient it can now be assumed that the semantic matching algorithm is able to rank the attractions based on the gathered interest profiles. Nonetheless, an ambient intelligence device computing individual tours might not be necessary, since the interests of the tourist are pretty much the same or fall into a couple of well-defined prototypical interest profiles. Therefore the next crucial question is: How diverse are the gathered interest profiles?

A way to assess the diversity independent of the actual distribution is to measure the entropy. Therefore each profile is interpreted as a combination of interests. Each combination has a certain probability of occurrence, which can be determined by dividing the frequency of each profile by the number of profiles in total. The single probabilities are then used to compute the entropy:

$$H = -\sum_{k=0}^{L} p_k * \log_2 p_k \quad \text{with L = number of profiles, p = probability of profile k} \quad (1)$$

The computed entropies are displayed in Table 3. The entropy values measure the average information content per profile in bits. If all profiles would be identical the entropy would be zero meaning that they do not contain any new information. If all profiles are different the entropy is the binary logarithm of the number of profiles. As the values are between 85% and 98% of the maximal entropy, most profiles do contain new information in comparison to the ones already provided. In other words, most interest profiles are different from each other to a certain degree and it might be difficult to split them into similar groups around a prototypical interest profile.

The entropy calculation for the interest profiles showed that the overwhelming number of profiles is different. However it might still be the case that there is considerable overlap between the interest profiles. Therefore each profile was compared against the other profiles, determining how many elements are identical. The average amount of identical elements is expressed as a percentage as listed in Table 3.

Table 3: Diversity of interest profiles: Entropy and Overlap

Method	Entropy	Overlap

	Max	Actual	Relative	Mean	Median	Min-Max
Tree	*6.23*	*5.76*	0.92	36%	53%	3%-53%
Images	*6.49*	*5.56*	0.85	24%	40%	1%-40%
Categories	*6.09*	*5.97*	0.98	27%	26%	17%-34%

The right side of Table 3 lists the overlap number. For the tree version 36% of overlap is given meaning that, on average, 36% of the classes in an interest profile are common with another interest profile. The relative entropy values indicated that the interest profiles are mostly unique and there is little overlap in interest elements between each other. Still the degree of difference between the interest profiles might make it difficult to split them into similar groups around a prototypical interest profile.

Furthermore the entropy of the distribution of selected interests within the ontological hierarchy can be determined. It gives an impression of whether the tourists select the same nodes within the same branch or if the selections are spread evenly across the whole tree structure. The absolute numbers of selections of each interest term were taken to calculate the entropy for each level of the hierarchy. What can be seen is that the interests are indeed very individual and tourists do not select the same things at all as the entropies reach more than 90%.

Table 4: Diversity of interest selection at various levels of the ontology illustrated for the image version

Image version	Entropy	Max Entropy	Relative
Level 1	2.28	2.32	0.98
Level 2	3.70	4.00	0.93
Level 3	3.98	4.17	0.95

4.5 Clustering

A basic value to express the degree of similarity between two elements is their distance. The aim is to determine the distances between the profiles to be able to make a statement about their similarity (Kaufman & Rousseeuw, 1987). The distance between two profiles depends on the distance of their elements which makes some definitions necessary:

1) $Dist1(e1, e2)$ → The distance between 2 interest elements (of one profile)

   It returns the number of branches between both interest elements within the ontology.

2) $Dist2(e, p)$ → The distance between an interest element and a profile

It returns the minimal value of distances between the single interest element and each interest element in the profile

$$Dist2(e, p) = MIN(\forall_{ie \ni p} : (Dist1(e, ie))) \quad (2)$$

3) $Dist3(p1, p2)$ → The distance between two profiles

It returns the maximal value of comparing profile 1 with profile 2 and profile 2 with profile 1

$$Dist3(p1, p2) = MAX\left(\frac{1}{p1.elements}\sum_{1}^{p1.elements} Dist2(element_i, p2), \frac{1}{p2.elements}\sum_{1}^{p2.elements} Dist2(element_i, p1)\right) \quad (3)$$

The distances for each profile to any other profiles are determined which results in a matrix.

Based on these distances the clustering was done by the following algorithm:

Foreach profile p1
Determine profile p2 with the lowest distance towards p1
If profile p2 belongs to a group
Add profile p1 to that group
Else
Create a new group with p1 and p2

Table 5: Grouping results

Method	# of groups	average # of profiles in group
Tree	*30*	2.5
Images	*40*	2.3
Categories	*31*	2.2

The algorithm creates pairs or groups of profiles, putting the closest related profiles together. The results indicate that the average number of profiles in such a cluster is small (Table 5). The clusters are mostly pairs.

For obtaining fewer and bigger groups the profiles need to be more categorical. But a number of 30 groups and higher with less than three typical profiles in it cannot be considered a cluster. As there are very few profiles being closely related two each other an individual interest elicitation appears to be indispensable.

5 Limitations and Implications for Future Research

The terms used to describe the different categories of interests might lead to misinterpretations and selections of other sights based on wrong associations. However, the reasonably good rank order correlations suggest that misinterpretations did not play a major role in the experiment. Further, the ranking of the sights by the tourists strongly depends on the quality of the pictures and text which are presented. The tourist might favor a sight regardless of his/her interests if it is displayed in a very nice shot and in contrast dislike a sight exactly matching his interests because of a less attractive picture. This could be a possible explanation for why the inspirational image method did not score higher.

So far the focus has been on the interactions of a single person with the DTG. In reality the percentage of tourists travelling alone is small compared to couples, families or groups. As proposed by Franke (2002) the concept of the DTG needs to be extended to serve groups of tourists jointly discovering a destination. Group management becomes necessary to synthesise interest profiles for groups from the individual ones, to support the common decision for the tour, to supply navigation instructions and interpretational information to all members and to enhance tour adaptation dealing with individual choices. Thus, an improved version of the DTG needs to allow for individual freedom as well as group experience.

6 Conclusion

In order to enjoy a destination to its full potential tourists need information about attractions that is tailored to their personal preferences. A Dynamic Tour Guide enables personalized and spontaneous guided tours. However, preference elicitation has to be particularly efficient in a mobile context. The findings presented in this paper suggest that offering a small selection of tour prototypes would not sufficiently capture the diversity of interest profiles among tourists. Rather, capturing preferences using a limited number of easy and intuitive attraction categories appears to be a suitable way in which preference data, necessary to make personalized recommendations, can be elicited.

The decision for the final design of the DTG will be a combination or an alternative of two methods: the main category version which is simple and the image version that offers to specify detailed interests that are actually existent as shown. Generally the whole tour specification process must be reduced to least inputs as possible. That means that beside the interests there isn’t much space for further inputs, as the figures showed that normal tourists are willing to spend only 2 minutes for the interaction in mobile context until they expect an appropriate result. That’s why the DTG will only present two more screens for defining the timeframe and for selecting a restaurant before the tour is created in a short time and presented to the tourist.

References

Abowd, G. D., Atkeson, C. G.; Hong, J.; Long, S.; Kooper, R.; Pinkerton, M. (1997): Cyberguide – a mobile context-aware tour guide, Baltzer/ACM Wireless Networks.

Cheverst, K.; Davies, N.; Mitchell, K.; Friday, A.; Efstratiou, C. (2000): Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences, Distributed Multimedia Research Group, Lancaster University, UK.

Franke, T. (2002). Integration of Group Support Tools into an Online Tourism Advising System. Workshop on User Modeling and Decision Making in Travel and Tourism Emergent Systems, Trento, April 9-10. Accessed online (September 6, 2005). Available at: http://ectrl.itc.it/home/umtts/program.htm

Freytag, T. (2003): „Städtetourismus in Heidleberg – Ergebnisbericht zur Gästebefragung 2003.“, Geographisches Institut der Universität Heidelberg.

Gretzel, U.; Fesenmaier, D.R. (2005): “Persuasiveness of Preference Elicitation Processes in Destination Recommendation Systems” Information and Communication Technologies in Tourism 2005, Andrew Frew et al. (eds.), Springer Computer Science.

Gretzel, U., Mitsche, N., Hwang, Y.-H., and D. R. Fesenmaier (2004). Tell me who you are and I tell you where to go – Travel personality testing for destination recommendation systems. Information and Communication Technologies in Tourism 2004. Vienna, Austria: Springer Verlag.

Kaufman L.; P. J. Rousseeuw (1987): “Finding Groups in Data: an Introduction to Cluster Analysis”. John Wiley & Sons, 1990

Kempermann, A.; Borgers, A.; Oppewal, H.; Timmermans, H. (2003): “Predicting the Duration of Theme Park Visitors’ Activities: An Ordered Logit Model Using Conjoint Choice Data” *Journal of Travel Research Vol.41*, Sage Publications.

Kempermann, A.; Chang-Hyeon, J.; Timmermans, H. (2004): “Comparing First-time and Repeat Visitors’ Activity Patterns in a Tourism Environment” *Consumer Psychology of Tourism, Hospitality and Leisure Vol.3*, CAB International.

Nguyen Q.N., Cavada D., Ricci F. On-Tour Interactive Travel Recommendations. In Information and Communication Technologies in Tourism 2004, Proceedings of the International Conference, Cairo, Egypt, January 26 - 28, 2004. Pages 259-270. Springer.

Ten Hagen, K.; Kramer, R.; Müller, P.; Schumann, B.; Hermkes, M. (2004): Semantic Matching and Heuristic Search for a Dynamic Tour Guide, Information and Communication Technologies in Tourism 2005, A. Frew et al. (eds.), Springer Computer Science

Acknowledgements

This project is part of VESUV in cooperation with Siemens AG, Microsoft’s European Innovation Center (EMIC) and Fraunhofer Institute IGD. VESUV is supported by the Federal Ministry for Economics and Labour (BMWA).