

# Off the beaten track - a mobile field study exploring the long tail of tourist recommendations

Nava Tintarev  
University of Aberdeen  
Dept. of Computing Science  
n.tintarev@abdn.ac.uk

Xavier Amatriain  
Telefónica Research  
Barcelona  
xamat@tid.es

Ana Flores  
Telefónica Research  
Valladolid  
anafc@tid.es@abdn.ac.uk

## ABSTRACT

This paper presents a field study of a framework for personalized mobile recommendations in the tourism domain, of sight-seeing Points of Interest (POI). We evaluate the effectiveness, satisfaction and divergence from popularity of a knowledge-based personalization strategy comparing it to recommending most popular sites. We found that participants visited more of the recommended POIs for lists with popular but non-personalized recommendations. In contrast, the personalized recommendations led participants to visit more POIs overall and visit places “off the beaten track”. The level of satisfaction between the two conditions was comparable and high, suggesting that our participants were just as happy with the rarer, “off the beaten track” recommendations and their overall experience. We conclude that personalized recommendations set tourists into a discovery mode with an increased chance for serendipitous findings, in particular for returning tourists.

## Categories and Subject Descriptors

H.5.2 [User Interfaces]: User-centered Design

## General Terms

Experimentation, Design, Human Factors

## Keywords

Recommender systems, user-centered design, field studies, mobile applications

## 1. INTRODUCTION

Recommender systems have been used in a number of domains to help users find items that are relevant to them, including books [14], movies [30], museum exhibits [15], news [22] as well as holiday destinations [24]. Recommender systems have the dual goals of improving user satisfaction with the items they consume while also increasing *benefits* to the catalog holder, or supplier of the items. To accomplish both, it is important to promote the exploration of “rare” (less well-known) items usually referred to as *the long tail* [12]. This paper explores how personalization can help improve tourist recommendations on mobile devices by promoting the long tail of tourist sites.

One of the challenges with the travel domain, is that common approaches using collaborative filtering techniques (or hybrids thereof) do not work particularly well: collaborative filtering techniques work best when there is a large user community, and many ratings for each user [20]. In contrast, travel activities are much less frequent than consumption in other domains such as books, movies and music. Another challenge, is the complexity of travel objects: we cannot simplify two trips to the extent that we can say that two travelers experienced the same trip despite certain similarities such as the destination. Also, if one simplifies travel description features like the destination, then this also damages the predictive ability of the recommender engine: already visited destinations do not offer sufficient information to predict the next destination.

Many previous approaches for travel recommendation have been on the level of destinations, or packages (e.g. hotel, flight and ski-pass). Others have been very knowledge intense. We consider to what extent recommendations can be used to personalize travel recommendation at the level of Points of Interest (POI) to recommend sights at a given location. We use a knowledge-based approach that leverages the existing knowledge on Wikipedia. The approach is generalizable and scalable, and requires very little input from the users to form the personalization.

Recommendations are also available from written travel guides and tourist offices. However, these often recommend popular or common sights to visitors with limited time to go sight-seeing. This means that the majority of travelers visit the same sights over and over again. In most cases these sights are worth visiting, but there is also no consideration for a given user’s particular taste. In this sense, the more rare, “off the beaten track” items get neglected and may never be recommended to anyone. A more complete recommender system should be able to have a higher coverage, and be able to cater to a variety of tastes.

Many studies discuss the hypothetical utility of recommendations, but very few measure their true utility. In this study we measure the actual effectiveness of the recommendations. We study whether the recommendations are followed in a mobile field study and how satisfied users are with the places they visit. Although the sample size is small, this is to the best of our knowledge the largest such field study evaluating recommendations to date. We now reiterate our research questions:

1. Does personalization lead participants to see more “rare” points of interest?
2. Do participants visit more of the points of interests

when their recommendations are personalized?

3. Are our participants more satisfied with the personalized lists?

This paper is structured as follows, first we introduce related work and our contribution in Section 2. Next, we go on give an overview of the system in which the recommendations were applied in Section 3. After this we look at the experiment where we evaluated the recommendations (Section 4) and the results of the evaluation (Section 5), followed by discussion and lessons learned (Section 6). Finally we conclude with our plans for future work in Section 7.

## 2. RELATED WORK

Current web-based tourist recommendations have different levels of granularity in the items they recommend. While the travel recommender system described by Waszkiewicz *et al.* [29] and TripSay [7] only focus on destinations (e.g. Moscow), systems such as the one described by McSherry [24], and Triplehop's TripMatcher [5, 19] recommend a complex vacation package including e.g. accommodation and ski-passes. What is lacking in these systems, though, is a user-specific trip consisting of (among other things) sights for this user to visit.

In this category, there are a number of systems which recommend, or at least present, a selection of POIs to visit in situ. The Cyberguide system, for instance, displays POIs on a map and offers context and location-aware specific information about nearby POIs [11]. Nevertheless, the system does not offer personalized recommendations.

Systems that *recommend* mobile tourist sights appear to require a great deal of knowledge elicitation for the domain. It takes a significant amount of effort to construct a complete ontology such as the one presented by Ardissono *et al.* [13]. However, this effort might result in reusable user models that may be extensible to other cities. Unfortunately, this model is not publicly available. In the COMPASS tourist application, the authors note that the general WASP platform can be augmented by domain specific prediction strategies [27]. They do not go into details of how to make the predictions for points of interest. The two latter approaches are flexible, neither offers replicable, low cost, and general approach for recommending for example sights in a city. However, this application was location-aware.

Similarly, while there are a number of ontologies for travel [6], and general POIs (including sights but also venues, hotels and restaurants) [16], none of them is applicable for classifying sight-seeing POI.

In addition, recent advances in technology now allow researchers to study factors that were previously not a possibility. There has been a move towards mobile field studies, considering how users behave "in the wild" [3, 18, 26]. In our study, participants are tourists who are able to learn about the POIs they are visiting using a mobile travel application, and we study their behavior during their touristic visit.

Another contribution of this work is to measure the effectiveness of recommendation: not only how many POI participants visited, but also how this relates to their self-reported preferences. Most studies of recommender systems, regardless of the domain discuss the hypothetical utility of recommendations, but few measure their true utility. It has been found that including a recommender system for video

on demand increasing the number of viewers, and that this number increased with the course of time [25]. Similarly, [21] report that the number of viewed and sold items (mobile games) increased during the usage of recommendations.

## 3. OVERVIEW OF THE CONSUMER SOCIAL GROUP TOOL



**Figure 1: General (non-personalized) query for a place ("Barcelona"), and results appear for three different countries: Spain, Venezuela and the Philippines.**

In the Consumer Social Group tool (CSG) a user can access travel related information, such as nearby POI, from their mobile phone during travel. The system provides recommendations based on their user profile. The functionalities used in the current prototype include the following.

1. **Search for places** (for example Barcelona). The user makes a place query. The framework returns a list of recommended places that match the query. In this case place is the name of a location such as Barcelona, and may require disambiguation. For example there are three places called "Barcelona" (see Figure 1), and a user searching for Barcelona will see all the options as a response to their query.
2. **Retrieve information about a place.** The user selects one place out of the places in the list. The system shows them the place information and also information for several POIs in the place as well.
3. **Search recommended POI.** The user makes a POI query and the system returns a list of recommended POIs that match the query (see Figure 4).
4. **Retrieve information about a POI.** The user selects a POI from the list of recommended POIs, and the CSG tool shows him the POI information (see Figure 5).

These functionalities are available both as desktop application via a browser and as a mobile tool for Android and

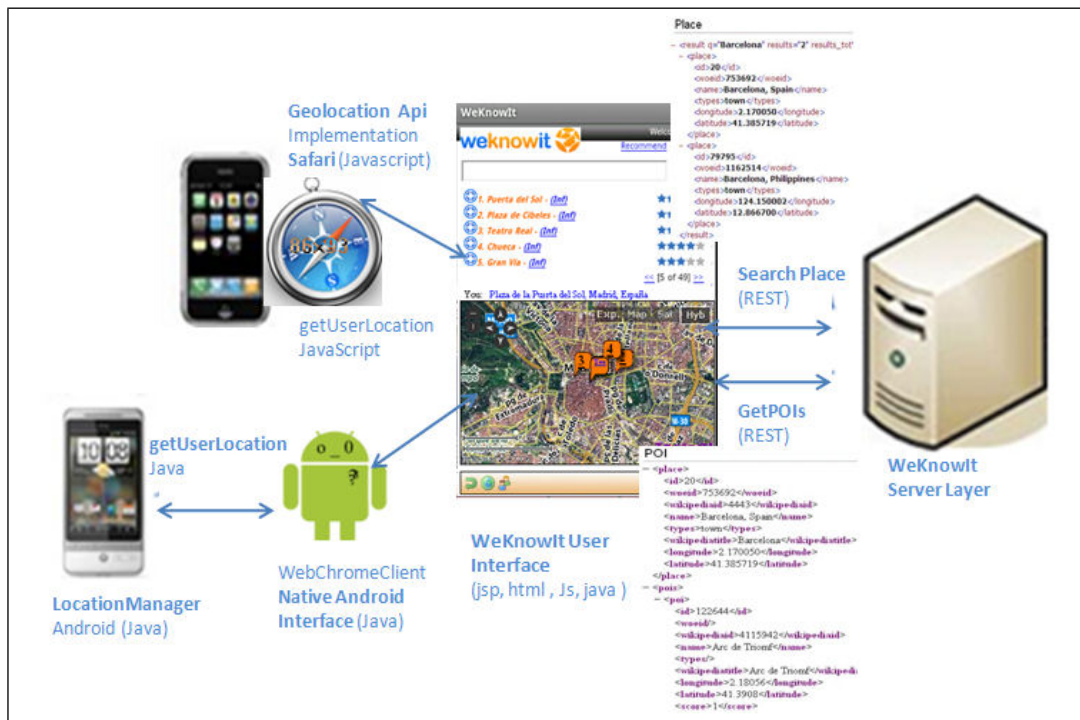


Figure 2: The architecture of the CSG mobile application

iPhone OS. We describe the mobile application in the following section.

### 3.1 CSG Mobile Tool

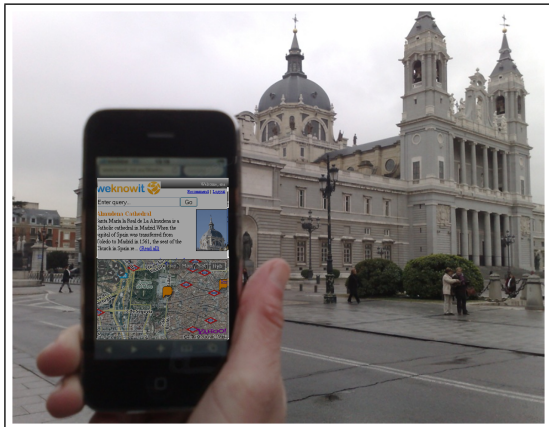


Figure 3: The CSG mobile application being used in Madrid, Spain

The mobile tool is a web-based adaption of the desktop CSG tool described in Section 3 and was developed for both Android and iPhone OS. It also includes other additional and different functionalities which take advantage of the knowledge of GPS position of the user to improve the system performance, and its usability. In particular, the mobile application allows the user to navigate the city while learning about POIs in situ.

Figure 2 shows the architecture of the application. When users log into the application, the mobile CSG application sets the user's current location as their starting point. To get the GPS location of the users, we have developed the interfaces for two platforms: one for iPhone OS and the other one for Android. The iPhone interface consists of a web interface uses an implementation of the Geolocation API (defined by W3C [8]), which is supported by version 4 of the Safari browser. For Android, we developed a Java native interface based on the WebChromeClient class of the Android API v1.5 [1] that makes use of the object LocationManager to get user's position. The Android native interface, and the web interface communicate via a JavaScript channel. The web interface is shared for both platforms, and is in charge of the communication with the web server to deploy the mobile CSG functionality.

When the system has the user position, it is also automatically and periodically processed by the system. The GPS location of users allow the mobile application to show the physical address corresponding to the GPS coordinates of the user, using the GClientGeocoder API for the reverse translation [2].

The users can also use the mobile application to find out more about POIs in the city, or request recommendations. When a user want to get his recommendations, they log onto the mobile CSG application, and press the link "recommend" on the homepage. Once users click on the link, the web interface makes a request using Ajax and REST to a recommendation service deployed on the web server. The recommendation service which retrieves the information applied for, returns an XML file with the POIS recommended and with an ordered list of exactly five POIs with scores in the range from 0-1, which are converted into a number of

stars from 1 to 5 in increments of 0.2. An example snippet of such an XML file is given below.

**XML 1** A portion of an XML file for the city of Barcelona.

```
<place>
<id>20</id>
<woeid>753692</woeid>
<wikipediaid>4443</wikipediaid>
<name>Barcelona, Catalonia, Spain, Catalonia</name>
<wikipediatitle>Barcelona</wikipediatitle>
<longitude>2.170050</longitude>
<latitude>41.385719</latitude>
</place>

<pois>
<poi>
<id>2405</id>
<woeid>-1</woeid>
<wikipediaid>59545</wikipediaid>
<name>Sagrada Familia</name>
<wikipediatitle>Sagrada Família</wikipediatitle>
<longitude>2.17444</longitude>
<latitude>41.4036</latitude>
<score>1</score>
<rank>1</rank>
</poi>
...
</pois>
```

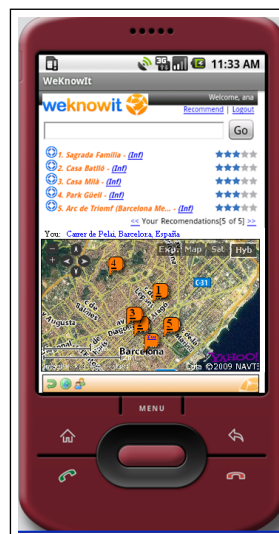
Finally, the recommended POIs are shown on the CSG mobile application as a list with the number of stars signifying the strength of the recommendation, as well as on a map so that the users can see their location relative to other POI (see also Figure 4). The scoring of POIs is described in Section 4.3, and was generated with the help of a call to an online service [10].

## 4. EXPERIMENTAL SETUP

The goal of this experiment was to see the effect of personalization on the behavior of participants. We wanted to know what kind of POI participants ended up seeing, if they were happy with their recommendations and which places they actually ended up visiting. In addition, we were interested in an investigation of the comments and behaviors that would arise in this type of field study.

### 4.1 Materials

Nine of the participants were supplied with HTC Android phones, and seven with iPhones on which they ran the CSG mobile tool. All phones had SIM cards with unlimited data plans so that they could freely use the browser to view the CSG tool. Two participants completed the experiment with paper lists of their recommendations, and regular maps, as there were an insufficient number of phones available. They had access the CSG tool prior to the study (as did the other participants) and conducted the experiment identically in other regards. We ensured that the mobile phones were started up and the batteries fully charged before the participants set out.



**Figure 4:** The top five recommendations for a given user in Barcelona, Spain, and displays their relative position on a map.



**Figure 5:** More information about a particular POI, in this case Casa Batlló, and the user's location relative to the POI.

## 4.2 Participants

Our participants were 21 members in a joint research project, and participated as an extension to a common meeting. The participants represented 8 countries: Austria, U.K., Cyprus, Czech Republic, Germany, Greece, Mexico, and Poland. Out of these participants 20 were male and 1 female, with an average age of 31.86 (StD=5.52).

We also surveyed the travel habits of our participants who stated the following reasons for traveling in the past: 28.6% mostly for work or business; 14.3% mostly for tourism or sightseeing; 57.1% a mix of work/business and tourism/sightseeing in equal measure, but on different trips (none of the participants specified that they travel for other reasons, or that they do not travel that much). The number of international trips made by the participants varied from 1 to 20 trips with a mean of 6.00 (StD=4.23).

Thirteen participants had been to Barcelona before, eight had not. This was reflected in the amount of knowledge the participants had about the city (9.5% said they knew absolutely nothing about the city; 33.3% said they had heard or read about *some* of the noteworthy sights; 5.3% said they had heard or read about *many* of them; 38.1% said they had *visited* some; and 14.3% said that they had *visited* many of the sights. Given the bias in this sample we tried to allocate participants who had been to the city previously equally between the two conditions described below (see also Section 6.1).

Participants were aware of the initial questionnaire, they had also completed parallel data collection tasks. They were not given access to any further detailed information on the personalization or experimental design. For example none of the participants knew if they were viewing the personalized or popularity list, or that there were two conditions that differed in this sense.

## 4.3 Generating POI Recommendations

Lists were either personalized or based on popularity, but both consisted of precisely five POIs given the limited time available for sightseeing. Both lists were based on the 273 POIs available for the city of Barcelona on Wikipedia [9]. We describe how the lists were constructed in the following sections.

### 4.3.1 Recommending Popularity

The “popularity” lists were all consisted of the same top five POIs. The ranking of these lists are based on *facets* as described in [28]. Facets are aspects or characteristics of POIs that were extracted by analyzing the relationships between two objects. When analyzing the relationship between two objects in Wikipedia, the authors define the *source* as the object to which the facet or feature belongs to, and the *target* as the object that represents the facet, and the type of facet *relation*. For example, given the objects “Bangalore India” and “Cubbon park”, the latter becomes a facet of the former, and the facet type is set to “subsumes”. Following this, facets were ranked for each POI. Ranking of facets is based on the statistical analysis of query terms (a) and query sessions (b) that are derived from image search logs. In addition, the tags (c) associated with the public photos in Flickr are used to complement the knowledge derived from the search logs. An aggregate ranking is derived based on a linear combination of the three sources (a,b and c). More information about ranking of facets can be found in [28].

### 4.3.2 Getting Personal

The *personalized lists* were based mainly on replies of a questionnaire. We asked participants for their gender, age and nationality. We also inquired how many international trips they have taken in the past year and whether they traveled more for business or for pleasure. In addition, we asked them about their previous experience and knowledge about the city of Barcelona. Finally, we asked participants to enter at least five keywords that best described what kind of sights they would like to see in the city. The input to this question was the main basis of the personalization, and was presented as a textbox with an auto-complete function. One example set of such keywords entered by a participant was “museums, cathedrals, parks, gaudí, monuments”.

The words that existed as auto-complete options were obtained as follows. First, we got the full list of POIs from the CSG tool (to make sure that the personalized lists are based on the same source as the popularity lists), and the Wikipedia categories for these POIs. Next, we removed all POIs containing the word metro because there are a large number of metro stations in Barcelona marked as POIs. While many of the metro stops are named after local landmarks, there is virtually no sightseeing value to any of them unless the tourist is particularly interested in transport systems. Similarly, we considered removing train stations from this list, but recognized that these often have architecture and historical value in and of themselves. Neighborhoods of the city were kept because these would be more difficult to extract automatically. We tokenized the full list into single words, and identified unique tokens. We also ensured that common stop-words such as “in” and “the” were not added to this list. We did not perform any stemming or other natural language processing.

POIs were then ranked according to the cosine similarity between a user’s keyword vector, and the keywords associated to each POI. In some cases ties would arise. For example, several items may similar to each other in terms of keywords, such as several art museums when the user has included the tags “art” and “museum”. These types of ties were resolved by considering the facet ranking of these items: the score associated to the POI was slightly incremented by the values derived from the popularity score.

In summary, we used a knowledge-based approach that leverages existing knowledge on Wikipedia. As such, the personalization does not require additional knowledge elicitation. The approach used here is generalizable and scalable, to the extent that the POIs are already identified and there are correct Wikipedia categories available for these POIs. In addition, the approaches requires very little input from the users to form the personalization - they only need to input five keywords. However, we note that the personalization is not meant to be the main contribution of the paper, and are well aware that using hybrid approaches may give additional benefits such as increased serendipity or the ability to improve recommendations over time.

## 4.4 The “Grand Travel Challenge” Field Study

Before starting the “Grand Travel Challenge” participants received a list of recommended points of interest. For half of these participants, the list was personalized, and for the other half it was not. All lists were static, i.e. the items on each list remained the same during the full course of the trip. In a pre-questionnaire all participants were asked to



	Personalized	Popular
Total number of POI visited	5.33 (1.67)	4 (1.41)
Number recommendations visited	2.08 (1.24)	2.78 (1.56)

**Table 1: Number of POI and recommendations visited. Reporting the average (StD) per list, in each condition**

specify if they had visited any of the POIs previously, and to say how much they think they would like each POI on a scale from 1 to 7 (1=not at all, 7=a lot). They were asked to rank the list as a whole (to consider intra-list factors such as diversity) (1=horrible, 7=great). They were also encouraged to elaborate on what they thought of the list and why.

Start and end location were identical for all of the participants, and these were clearly specified well in advance as well as reiterated on the day of the Challenge. Participants were told that they were allowed to travel freely for five hours using a mobile phone. They were also told that they were not required to see any particular number of POI. Each participant was explicitly told that these were only recommendations and that they were free to visit whichever places they wanted during this time. In addition, they were instructed to travel alone. To increase the likelihood of individual travel participants were dispersed from a central starting location at three minute intervals. To their aid, each participant was given a “survival pack” containing a mobile phone, a map, a metro card, and some smaller items such as a bottle of water.

To ensure that participants were actively sight-seeing they were asked to take photos near the landmarks that they visited, and at least one photo at each sight. The instructions were not to necessarily take the aesthetically appealing, or most representative of the location, but the most interesting.

Participants were asked to present their photos with a story for the actual “Challenge” that took place in the afternoon of the same day. Participants voted for the best photo with one vote per participant, and the winner received a prize.

Finally, post-questionnaires were collect to analyze the effectiveness of, and satisfaction with the recommended POIs. In the post-questionnaire we asked participants to tell us which of the recommended POIs they had visited, and how much they had enjoyed their visit to each of them. We also asked them to list POIs they had visited that were not in the list. As before, participants were able to rate the overall list, and elaborate on the reasons behind their ratings.

## 5. RESULTS

In this section, we report on the different results gathered from both pre- and post-questionnaires. Given the small sample sizes we have elected not to conduct any statistical tests to compare differences between conditions. While it is arguable that non-parametric tests might be used for smaller samples, we believe it prudent to report our results as trends.

### 5.1 Effectiveness

One of the things we wanted to know was whether the recommendations were truly effective, i.e. were the recom-

	Personalized	Popular
Number of popular sights recommended	0.42 (0.51)	5 (0.00)
Number of popular sights visited	1.25 (0.87)	2.78 (1.56)
Popularity score	0.14 (0.23)	0.39 (0.32)

**Table 2: Popularity of POIs. Reporting the average (StD) per list, in each condition**

	Personalized	Popular
Number of POI visited previously	0.67 (0.98)	2.33 (2.24)
Novelty for participants that have been to the city previously	0.91 (0.13)	0.64 (0.41)

**Table 3: Average novelty for visited POIs per participant, in each condition**

mendations any good? Firstly, we wanted to know if the lists we gave to participants were too short or too long. We see in Table 1 that the average number of places visited per participant is close to the list size of five for both conditions. Participants on average had time to see five POI, but did not exceed this number greatly. We note however that participants *saw more sights overall in the personalized condition*.

Given that participants were not bound to the recommendations and free to go wherever they pleased, we were also curious to see how many of the recommended POI were visited. In the Table we see that the number of recommendations that participants followed is larger for the popular recommendations. This is contrary to our initial expectations, that participants who received personalized recommendations would visit more of the recommended POIs than those who received popular recommendations. The discussions in Sections 6.2 and 6.1 consider possible reasons for this finding.

### 5.2 Popularity

We saw that although the participants with personalized lists did not follow all of the recommendations, they did go see many POIs. The places that participants went to, include those discovered on their own, and were rarely one of the “top five” recommendations. We were curious how much more the visited places were “off the beaten track”. Table 2 summarizes the results. The score of an item ranges from 0-1 (inclusive), and is based on a Yahoo! API service [10]. It is worth noting that the score drops very quickly with only 118 out of 273 items having a non-zero score. The first most popular POI has a score of 1, while the next item already drops to 0.52.

We saw that the participants in the popular condition followed more of their recommendations, and that they had been to more of their recommended POIs *on a previous visit* than the participants who received personalized recommendations (Table 2). So, we wanted to find out if they just came back to their favorite popular destinations, or if they discovered as many novel items as the participants in the personalized condition.

We define *novelty* as the number of new places a participant visited, over the total number of places visited. Here

	Personalized	Popular
User rating for visited and recommended POI	5.00 (1.72)	5.88 (1.36)
User rating visited but not recommended POI	5.97 (1.00)	6.00 (1.15)
User rating all visited POI	5.58 (1.40)	5.97 (1.29)
Mean satisfaction with overall list (pre-questionnaire)	4.25 (1.14)	5.44 (1.33)
Mean satisfaction with overall list (post-questionnaire)	4.42 (1.24)	5.11 (1.62)

**Table 4: Satisfaction with the visited POI (after) the challenge, in each condition**

	Personalized	Popular
Cumulative satisfaction	28.92(10.18)	22.33(9.99)

**Table 5: Cumulative satisfaction is the sum of reported satisfaction on all visited POIs**

by “new” we refer to locations participants has never seen in the past (including recommendations, and self-discoveries). In Table 3 we see that personalized lists indeed led our returning tourists to visit more new places than the popular lists.

### 5.3 Satisfaction

With regard to satisfaction we consider a number of variations in Table 4. In the Table we see that for the POI that are both recommended and visited, the *average satisfaction is high*. The score is slightly lower for the personalized condition, but is roughly and comparable between conditions. A similar result is found for POI that participants found on their own, as well as the mean satisfaction across all visited POI. That is, while people in the personalized condition are viewing more rare POIs, they are still generally happy with the POIs they visited – and on average rate the recommendations as 5 or 6 on a 7 point scale (7 being = “I like it a lot”). Interestingly, participants were even happier with the POIs they discovered on their own in both conditions than the recommendation they received.

The score given to the list as a whole is a list-wide metric, and considers items that were not visited as well as additional factors such as e.g. diversity. The *overall list score is lower for the personalized condition*, but the score for the popular list decreases from the pre- to post-questionnaire. We discuss the change in satisfaction further in Sections 6.1 and 6.3.

#### 5.3.1 Cumulative Satisfaction

Summary statistics included in Table 5.3 and discussed above do not capture all possible interpretations of the overall satisfaction of a tourist on a sightseeing tour. If a tourist is on a sightseeing tour with limited time (e.g. one day), the number of places visited in that period might have a positive impact on the overall travel experience, beyond what the average rating of all visited places is. In particular, note that the average satisfaction that we report is penalizing ad-

	Personalized	Popular
Average $\Delta$	0.21 (1.89)	0.14 (1.39)
Average $ \Delta $	0.90 (1.04)	1.29(1.37)

**Table 6:  $\Delta$  is the difference between the pre-rating (R1) and post-rating (R2): R1-R2.  $|\Delta|$  is the absolute value which shows the absolute value of the difference.**

venturous strategies in which you might need to visit several below average POIs before discovering an excellent one.

We therefore define the *Cumulative Satisfaction* as the sum of the reported ratings for each individual POI visited. Results for both conditions are reported in Table 5. We see that the value is higher for personalized recommendations than for popularity lists. We conclude that although average satisfaction might be lower in the personalized case, the fact that we are increasing the number of visited places might result in a higher satisfaction on the overall experience.

### 5.4 Change in opinion

**Figure 6: Signed delta for the two conditions.**

We also looked at the change in opinion for the items that people had visited. That is, we looked at the difference between the pre-rating (R1) and post-rating (R2): R1-R2 (see Table 6). A positive value denotes an overestimation, and a negative value an underestimation. We found that the signed average was similar for the two conditions, but that the absolute value was higher for the personalized condition. People get their estimations wrong more often with personalized recommendations (high StD), but also that these estimations can be over and underestimations in similar measure (as these seem to cancel each other for the signed difference).

In Figure 6, we analyze the details by looking at how the differences in pre-post opinions are distributed. We see that values are much more spread for users with the personalized lists. This means that users getting the popular recommendations were much more aware of what they would be seeing before actually visiting the POI. Of course this could have a negative interpretation if users with personalized lists were always overestimating how much they would like a place. This is not the case, underestimations are around 40% in the personalized condition vs. less than 30% in the non-personalized. On the negative side, though, extreme overestimations of +4 are slightly larger in the personalized case.

We conclude that our personalized recommendations are greatly increasing serendipity (*i.e.* chance of visiting a place that we liked much more than we thought) at the risk of slightly increasing the possibility of deceiving expectations for some recommendations.

## 6. DISCUSSION

In this section we discuss our findings and propose explanations for our results.

### 6.1 Post-hoc analysis

After our initial analysis we found that the number of participants that had been to the city previously was larger in the personalized condition, as we can see in Table 7. We had

	Personalized	Popular
Been	8	5
Not been	4	4

**Table 7: Number of participants who had been to the city of Barcelona in the two conditions**

distributed participants who had been to the city previous equally between the two conditions initially. However, this balance was altered by the fact that a few of the intended participants could not use their phones overseas or had not brought phones to the study as requested. Naturally, the fact that the number of participants that had been to the city before was unbalanced, could have effected the results reported in this paper. For this reason, we reran the analysis looking only at participants who had been in the city before. The trends are in the same direction as before with the following considerations.

The average number of recommended places visited is slightly *lower* for the popular lists than personalized when we are looking only at participants who have been in the city before. It was higher when we looked at all the data. This suggests that participants do not go back to the popular places again and again. Rather, the participants in the popular condition that went to the popular places were the ones that had not been to the city before. Similarly, the average number of popular places visited drops for participants who received the popular list when the analysis only considered participants that had been in the city before.

There was also a change for the satisfaction with recommendation list as a whole if we only look at returning visitors. Satisfaction drops between the pre- and post-questionnaires for both conditions, rather than just for the popular condition. However, the drop is greater for the popular list.

## 6.2 Effectiveness

We saw that participants who received the popular list visited more of their recommendations than those who received personalized lists. However, we saw in our post-hoc analysis that this trend is reversed if we only look at participants who had been to the city before.

This suggests that new visitors to the city may have been influenced by factors such as the availability of the popular POIs - the names of popular sights were easier to recognize. In comparison, first time visitors might not have been able to recognize many of the POIs in the personalized condition, nor have been able to recall many of the popular sights. These participants might not have trusted the recommendations they were given as much as in the those in the popular condition. Returning visitors on the other hand, were likely to be familiar with names, but not be as keen to visit popular sights as they had seen them before. The popular lists are thus more effective, but this seems to be more the case for first time, rather than returning, visitors.

It is hard to differentiate the effects of low trust and the effects of effectiveness, which both would lead to fewer followed recommendations. What is worse, the effects of trust and availability would have influenced the initial satisfaction with the lists as well. However, satisfaction *after* visiting the sights offers a complementary evaluation of the effectiveness of the lists that may be more genuine than how many rec-

ommended POI were visited.

## 6.3 Satisfaction

We give two possible explanations why the overall satisfaction with the recommendation lists (in both conditions) are lower than the score for the POI that were visited. Either the participants filtered out the “poor” recommendations from the very beginning, and refrained from visiting them, or the personalized list was strong, but our participants filtered out options they had not heard of and were unsure about<sup>1</sup>. We note however that the *cumulative* satisfaction, that considers the number of visited POI as well as the score they received is higher for personalized lists.

## 6.4 Popularity

It is arguable that the personalized list led to participants seeing more unpopular items, just as the popular lists led participants to see popular items. However, we argue that these lists put participants in the personalized condition in “discovery mode” where they found more rare POI. In support of this argument we recall that the average participant in this condition visited less than *half* of the POI that were recommended to them. This suggests that the low popularity value for the items seen by participants who received the personalized lists is not only due to the properties of these personalized lists, but the properties of *all* the POIs that the participants in this condition visited. The fact that participants in the personalized condition saw *more* POIs overall also supports the idea that this condition facilitated a “discovery mode”. While our participants were happy with being recommended the same popular sights again, these recommendations lack in novelty: in the long run, a recommender system would not be able to recommend the same top sights over and over again to the same people. The discovery mode triggered by the personalized lists, help these travelers to discover new places, and maintain their satisfaction with the travel experience over time.

## 6.5 Qualitative comments

We noticed that participants often noted contextual factors for not visiting certain POIs. For example some participants mentioned that they did not visit museums as there was good weather that day: “*since the weather was good I avoided inside places (e.g. the two museums)*”. We noticed that such explanations were given exclusively for personalized recommendation lists. While explanations like these are absolutely reasonable, and while context such as weather are important in making travel recommendations, it is interesting to note that there were no such problems for popular recommendations, such as visiting Casa Batlló and Casa Milà, both of which are indoors.

## 6.6 Lessons learned for mobile field studies

Mobile field studies are not without difficulties [17]. While we had access to a sizable number of tourist participants, the number was not large enough to be able to test more types of personalization or exclude participants that did not fulfill certain criteria such as not having visited the city of

<sup>1</sup>Participants were able to use the CSG tool during their trip to look up the POI, but this information was occasionally missing for less popular POI. Thus in these cases participants did not have much information the recommended POIs



Barcelona before. Thus, while the method of personalization is not particularly sophisticated, we aimed to use an approach that is generalizable, replicable and has an accepted precedent in the literature. We do not claim to have developed a new method of personalization as such, but merely evaluate an commonly accepted approach.

Other difficulties are more technical in nature, such as the drain of GPS on battery life. In our case we had hoped to record a trace of participants movements via GPS, which the participants could choose to include in analysis. Given the drain on battery life however, some of the traces were incomplete. The remaining traces were thus rendered unworkable for analysis. Failing that, we also considered using the geo-tagged photos that participants took during the challenge and were asked to upload to Flickr.com. Also here, we encountered a number of technical difficulties uploading the images and ensuring they belong to the same account or group, such as the upper quota on free accounts. As such we encourage future field studies to always consider backup solutions when (not if) difficulties occur.

## 7. CONCLUSIONS AND FUTURE WORK

Although participants did not follow many of the recommendations in the personalized lists, we found that personalized recommendations enabled a “discovery mode”. In this mode, participants visited more POIs than in the popular condition, and these POIs were also more rare than POIs visited by participants in the popular condition. The level of satisfaction between the two conditions was comparable and high, suggesting that our participants were almost as happy with the more rare, “off the beaten track” recommendations that they received. If we consider a cumulative model for satisfaction, in which the number of visited places also contributes positively, we see that this increases for personalized lists.

We also saw that the number of followed recommendations was comparable between conditions, for participants that had been in the city before. It would seem that returning visitors that saw the popular lists, go “nowhere much”. They are happy with their recommendations (slightly happier than personalized), but are actually missing out. Returning visitors in the popular condition did not discover as many novel items as the participants in the personalized condition. Finally, we have also shown that personalized recommendations may increase serendipity, since users are more likely to discover sites that surpass their a priori assessment. The discovery mode that the personalized lists lead to help these travelers to discover new places, and maintain their satisfaction with the travel experience over time. We caution however that given that many of our participants had previously been to the city of Barcelona, our reported trends are more likely to be applicable to tourists who are already somewhat familiar with a city.

The practical implications of our findings are the following. *Popular points of interest may be suitable recommendations for first time tourists with limited time. However, popular recommendations also limit the serendipity of the overall experience, and may cause travelers to be less adventurous. In particular, for returning visitors, more rare POIs help maintain satisfaction and interest in the place or city they are visiting over many trips.*

In this experiment, participants were asked to travel and navigate alone. While this happens occasionally in a natural

scenario, traveling is very often a group activity and it is our intention to consider aiding the group as well as individuals. For example we may imagine that this could influence the scoring of recommendations, as been investigated in other work such as e.g. [23]. This work only considers the initial prototype of the CSG application, but we are already working on a functionality that allows to view the location of friends.

Other plans for future additions include recommending itineraries: recommendations will no longer be single points of interests, but itineraries with several POIs. As such, itineraries will not only consider the user’s current location, and the user’s interests, but also how much time they have to spend on sightseeing, and travel times between POI locations given different modalities (e.g. walking or car). We are currently considering the applicability of such a service combined with augmented reality for the Olympics 2012 in London, while augmenting the POI to include venues and events.

## 8. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community’s Seventh Framework Programme FP7/2007-2013 under grant agreement n°215453 - WeKnowIt. Many thanks to Ana Flores and Manuel Vicente for the development of the CSG tool. This paper also made use of the SimMetrics Java package[4] developed by Sam Chapman.

## 9. REFERENCES

- [1] Android location manager. Last retrieved April 2010. <http://developer.android.com/reference/packages.html/>.
- [2] Geocoder. Last retrieved April 2010. <http://code.google.com/apis/maps/documentation/services.html>.
- [3] Reality mining at MIT. Last retrieved April 2010. <http://reality.media.mit.edu/>.
- [4] Simmetrics. Last retrieved April 2010. <http://www.dcs.shef.ac.uk/~sam/stringmetrics.html>.
- [5] Ski-europe. Last retrieved April 2010. <http://www.ski-europe.com>.
- [6] Travel ontology. Last retrieved April 2010. <http://www.schemaweb.info/schema/SchemaDetails.aspx?id=236>.
- [7] Tripsay. Last retrieved April 2010. <http://www.tripsay.com>.
- [8] W3c geolocation api. Last retrieved April 2010. <http://dev.w3.org/geo/api/spec-source.html>.
- [9] Wikipedia. Last retrieved Dec. 2009. <http://en.wikipedia.org/>.
- [10] Yahoo API for getting POIs. Last retrieved Dec. 2009. <http://weknowit.research.yahoo.com/weknowit/apis/getPOIs.php?id=2405>.
- [11] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide: A mobile context guide. In *ACM Wireless Networks*, pages 421–433, 1997.
- [12] C. Anderson. *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion, 2006.
- [13] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso. Intrigue: Personalized recommendation of

- tourist attractions for desktop and handset devices. In *Applied Artificial Intelligence*, 2003.
- [14] M. Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Proceedings of the Workshop Beyond Personalization, in conjunction with the International Conference on Intelligent User Interfaces*, pages 13–18, 2005.
- [15] F. Bohnert and I. Zukerman. Non-intrusive personalisation of the museum experience. In *UMAP*, pages 197–209, 2009.
- [16] M. Braun and A. Scherp. Collaborative poi. Last retrieved April 2010. <http://isweb.uni-koblenz.de/Research/systeme/csxPOI>.
- [17] K. Cheverest, N. Davies, K. Mitchel, A. Friday, and C. Efstratiou. Developing a context-aware electronic tourist guide: Some issues and experiences. In *CHI Letters*, 2000.
- [18] K. Church, J. Neumann, M. Cherubin, and N. Oliver. Socialsearchbrowser: A novel mobile search and information discovery tool. In *In Proceedings of the International Conference on Intelligent User Interfaces (IUI)*, 2010.
- [19] J. Delgado and R. Davidson. Knowledge bases and user profiling in travel and hospitality recommender systems. In *Proceedings of the ENTER 2002 Conference*, pages pp. 1–16, Innsbruck, Austria, January 22–25 2002. Springer Verlag.
- [20] A. Felfernig, S. Gordea, D. Jannach, E. Teppan, and M. Zanker. A short survey of recommendation technologies in travel and tourism. *ÓGAI Journal*, 25:17–22, 2007.
- [21] D. Jannach and K. Hegelich. A case study on the effectiveness of recommendations in the mobile internet. In *Conference on Recommender systems*, pages 205–208, 2009.
- [22] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. GroupLens: Applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3):77–87, Mar. 1997.
- [23] J. Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. *User Modeling and User Adapted Interaction*, 14:37–85, 2004.
- [24] D. McSherry. Explanation in recommender systems. *Artificial Intelligence Review*, 24(2):179 – 197, 2005.
- [25] M. Perrero, F. Antonelli, and M. Geymonat. Recommendation on tv: What do users want? a user study. In *Recommender-based Industrial Applications Workshop in association with the conference on Recommender Systems*, 2009.
- [26] D. Rowland, M. F. L. Oppermann, J. Marshall, A. C. and Boriana Koleva, S. Benford, and C. Perez. Ubiquitous computing: Designing interactive experiences for cyclists. In *MobileHCI*, 2009.
- [27] M. van Setten, S. Pokraev, and J. Koolwaaaji. Context-aware recommendations in the mobile tourist application compass. In *Adaptive Hypermedia*, pages 235–244, 2004.
- [28] R. van Zwol and B. Sigurbjörnsson. Faceted exploration of image search results. In *WWW*, 2010.
- [29] Waszkiewicz, Cunningham, and Byrne. Case-based personal travel agent. In *International Conference on User Modeling*, 1999.
- [30] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW'05*, 2005.

## APPENDIX

### A. PERSONALIZATION QUESTIONNAIRE

The personalization questionnaire is available online: <http://78.46.87.99/tourist/>.

### B. PRE-QUESTIONNAIRE

An example pre-questionnaire is available online: <http://78.46.87.99/tourist/preQnaire.doc>. This is an example of a personalized recommendation list for the keywords “architecture, history, botanical, music, archaeological”.

### C. POST-QUESTIONNAIRE

The post-questionnaire is available online, and corresponds to the above personalized pre-questionnaire: <http://78.46.87.99/tourist/postQnaire.doc>.