# Understanding People's Place Naming Preferences in Location Sharing

Jialiu Lin, Jason Hong, Norman Sadeh

June 2009
CMU-ISR-09-121
CMU-CyLab-09-010

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This work is supported by NSF Cyber Trust grant CNS-0627513 and ARO research grant DAAD19-02-1-0389 to Carnegie Mellon University's CyLab. Additional support has been provided by Microsoft through the Carnegie Mellon Center for Computational Thinking, FCT through the CMU/Portugal Information and Communication Technologies Institute, and through grants from FranceTelecom and Nokia. The authors would also like to thank Erick Fang, Michael Benisch, Patrick Gage Kelley, Janice Tsai, Eran Toch, and Jackie Xiao for their assistance with our study.

**Keywords:** Location sharing, Location-based service, Location representation, Place naming.

## ABSTRACT

Many existing location sharing applications provide coordinate-based location estimates and display them on a map. However, people use a rich variety of terms to convey their location to others, such as “home,” “Starbucks,” or even “the bus stop near my apartment.” Our long-term goal is to create a system that can automatically generate useful place names based on real-time context. Towards this end, we present the results of a week-long study with 30 participants to understand people’s preferences for place naming. We propose a hierarchical classification on place naming methods. We further conclude that people’s place naming preferences are complex and dynamic, but fairly predictable using machine learning techniques. Two factors influence the way people name a place: their routines and their willingness to share location information. The new findings provide important implications to location sharing applications and other location based services.

# 1 INTRODUCTION

Over the past few years there have been a growing number of “friend finder” applications which let people share their location information with others[1] [2] [3][16, 29, 36]. These friend finders typically provide coordinate-based location estimates and show people’s locations on a map. These visualizations are a good match for navigation and emergency response applications which require absolute locations. However, these displays often lack semantically useful names for the people viewing these displays. People usually do not describe their locations to others as, for example, “40.443 north, 79.941 west,” or even “5000 Forbes Avenue.” Instead, people tend to use meaningful names such as “home,” “Starbucks,” or “near Liberty Bridge.”

These kinds of place names can be useful in integrating location information with other services. For example, one could show his or her current location as a status message in instant messaging clients, or show a text label denoting the place a photo was taken in a photo sharing application. Appropriate place names can also be useful in preserving a user’s privacy while still providing utility. For example, a person might be willing to share they are at “home”, but not the street address or geo-coordinates of their residence.

However, there is a gap between how people describe places and what technology can currently offer [41]. Using existing reverse-geocoding systems, it is possible to translate geo-coordinates into street addresses, neighborhoods, postal codes, and countries. However, it is not clear how useful these kinds of names would be in everyday use. In fact, based on our studies presented in this report, people rarely, if ever, use these kinds of names to describe their locations. If we are to build a system that can generate useful semantic names for places, we need to have a better understanding of what factors influence how people name places when sharing location information with others.

As a first step towards this end, we present the results of a study with 30 participants examining preferences for place naming, that is, how people name places. We recorded the location traces of our participants over a week, and followed up with participants specifying place naming preferences. By analyzing all the place names collected in our study, we were able to identify several general place naming patterns based on different kinds of information implied. We analyzed the influence of various attributes in how people refer to a place. We found that:

- People have very complex and dynamic place naming preferences. However, by applying machine learning techniques, we can predict the place naming methods people used with an average accuracy larger than 90%.
- Two factors that significantly influence the choice of place naming methods are whether the place is a routine place, and how willing the person is to disclose this location information to others.

[1] Helio, http://www.helio.com
[2] Locaccino: A User-Controllable Location-Sharing Tool, http://www.locaccino.org
[3] Loopt, http://loopt.com

- People are more open to sharing their location information in the form of place names instead of exact positions.

The remainder of this report is organized as follows: In Section 2, we present previous relevant work and its relations to our study. In Section 3, we describe the procedures of our user study and explain the implementation of the tools used in our study. In Section 4, data analysis will be presented. In Section 5, we will discuss the crucial findings in our user study and their implications to existing location sharing applications. Finally, we present the conclusions and our future work direction in Section 6.

# 2 RELATED WORK

In the 1970s, researchers in social interactions and environmental psychology domains showed several underlying meanings of locations[21, 31, 37]. A meaningful place name can capture the location's demographic, environmental, historic, personal, and commercial significance[13]. Incorporated with other knowledge, location information can also be used to infer higher level contextual information, such as people's activities, availability, interruptibility, and safety (for example [11, 19, 26, 35, 38]). This latter possibility of inferring higher level context has been a very active area of research.

One important observation on place names is that one person can associate multiple place names to the same place under different conditions. For example, a person might refer to her work place as "office" to someone who knows that place. But she might use the exact building name and room number to refer to the same location if the recipient of this information has no prior knowledge of that place. In 2005, Zhou el al.[42] pointed out this dynamic feature of place names and investigated the types of descriptions people naturally produce for places. Their work also touched slightly on the factors that influence the place naming in a qualitative manner.

Due to the dynamic nature, how to automatically generate appropriate place names remains a research challenge. To the best of our knowledge, little work has been done in generating place names according to different contexts. However, there are several directions closely related to place naming. Here we briefly summarize the literature related, location sharing applications, place discovery and place labeling.

## 2.1 Location Sharing Applications

With the rapid development of the positioning technology and the large demand for high-end mobile devices such as smart phones, estimating one's location becomes much easier. During the past few years, systems that provide location sensing and sharing services have been attracting lots of interest both from industry and academia[1] [3] [4] [5][16, 29, 36].

Researchers found that people have significant privacy concerns when sharing their location with others[4, 5, 9, 15, 17, 22, 30]. For example, in a study designed to determine the willingness of users to share their current locations with people in their social network, Consolvo et al.

---

[4] Google Latitude, http://www.geogle.com/mobile/default/latitude.html
[5] PlaceLab, http://www.planet-lab.org

observed that the information disclosed depended primarily on the relationship between the requester and requestee, the user's current activity, and both the user and the other person's current location[8]. Iachello et al. identified the use of four different deception techniques used for inaccurate disclosure: delayed response, time shifted response, ignored response, and explicitly deceptive[17]. Their conclusions suggested that it is essential for the location sharing application to support plausible deniability to disclosure location information of the user. Cornwell et al. developed and evaluated a people finder application which demonstrated the user-controllable security and privacy mechanisms[9]. Through a series of user studies, researchers had captured the users' privacy policies in the setting of location sharing. Following this trend, some recent applications provide users more control of their privacy preferences[2][30, 33]. Locaccino [2] is a user-controllable location sharing tool which gives users precision control on selectively sharing their location. Users can specify privacy polices not only on who can view their locations but also can create rules based on temporal and spatial restrictions.

It is clear that privacy control mechanisms are necessary for location sharing. Our focus is to understand people's preferences for place naming which we believe are more complex and dynamic than people's privacy preferences.

## 2.2 Place Discovery

There has also been a fair amount of work in using traces of people's locations to extract places. Place discovery algorithms also seek to bridge the gap between geo-coordinates and places [10, 40], by finding the boundaries of semantically meaningful places rather than giving names to them. How to extract significant places or recognize users' behaviors according to their location histories are hot issues in machine learning [1, 2, 23-25].

Ashbrook et al. extracted significant locations by clustering GPS data taken over periods of time at multiple scales [1]. Their work can also be extended to a multi-user scenario [2]. Similarly, Liao et al. successfully extracted people's activities and significant places from traces of GPS data. Their system used hierarchically structured conditional random fields to generate a consistent model of person's activities and places [23, 24]. Based on clustering, users' movement can also be predicted by using a Markov model. Along similar lines, Zhou et al. built a place discovery system based on users' location data and evaluated their system by comparing the discovery results with the ground truth captured in user interviews [39, 40]. Hightower et al. used the WiFi and GSM radio fingerprints to automatically learn the places in order to recognize them when they were visited again[14]. Predestination [20] by Krumm et al. uses the history of a driver's destinations, along with data about driving behaviors, to predict where the driver is going as a trip progresses.

In general, this past work focused on using location traces to identify areas that are significant to people. However, they didn't come up a way to automatically assign names to these recognized places. In contrast, our work is focused on associating meaningful names and other information with these extracted areas. The work in this report focuses specifically on a user study to understand how people associate names with places, as part of a larger goal of creating a system to support this activity.

### 2.3 Grassroots Place Labeling

An alternative way to obtain place names is to aggregate place names from grassroots contributors[13, 27].

Websites such as Wikimapia[6] and Flickr[7] encourage users to tag their resources. A significant number of labels can be generated by aggregating the labels contributed by other people. For example, Rattenbury et al. proposed an approach for extracting and distinguishing “place” and “event” semantics from tags, unstructured text-labels assigned to resources on Flickr based on each tag’s usage patterns [28]. However, these methods also face several problems such as how to eliminate “bad” labels, how to create incentive for users to contribute, and how to preserve contributors’ privacy.

Wang et al. proposed four different prototypes of place annotation system on mobile phones and compared their usability through a series of user studies[34]. Their findings suggested great implications on how to make a place annotation system more useful. However, their method still relied on human labeling and didn’t solve the fundamental problem.

Another crucial drawback of human labeling is that they cannot capture the dynamic nature of place names. Thus, they may not be the optimal toward automatically naming places.

### 2.4 Computing Models for Places

Schilit et al. proposed a hierarchical location model to index different locations within a certain region and at different granularities [32], such as regions, buildings, and floors. Similarly, Jiang et al [18] proposed a computable location identifier that used a URL-like string to define the hierarchical structure of different locations. This method worked well on representing locations’ physical affiliation. However, it only captured partial semantic meanings of these locations, and thus was less useful to human beings. What’s more, this method was very difficult to scale up due to the tremendous efforts needed for defining the hierarchical structure beforehand.

## 3 AN EMPIRICAL STUDY OF PLACE NAMING

In this section, we discuss two studies that we conducted to gather real-world data reflecting users’ preferences of naming places under different situations. The first study was an informal pilot study to help us understand various parameters of this space. The second was a week-long user study that included getting traces of people’s locations and understanding how they named those places.

---

[6] Wikimapia, http://www.wikimapia.org
[7] Flickr, http://www.flickr.com

## 3.1 Experiment Overview

### *3.1.1 Pilot Study*

We conducted a pilot study with 5 participants (2 males and 3 females) in early February 2009. Each participant was asked to list several places they were at in the past week and assign names that she would use to refer to these places to different groups of people in her social networks. Participants were also asked to identify some factors that influenced the way they named those places.

Based on their answers, we noted that there are many ways people used to refer to a location, including using some generic terms (e.g. ‘home,’ ‘work’), using part of an address (e.g. ‘Beacon Street,’ ‘CIC building’), using the function or business nature of the place (e.g. ‘bank,’ ‘Starbucks’), as well as using nearby landmarks known to people that live in the area (e.g. ‘near Liberty Bridge’). We also observed that people sometimes use a combination of more than one of the above methods, such as “Starbucks on Craig Street” (business + address), or “Restaurant near campus” (function + generic). These observations gave us some ideas of what resources might be useful in generating meaningful place names.

We did not see highly specific and contextualized names like “near the bench where we first met”. While these kinds of names could be useful in certain situations, we believe that they are less useful in the more general location-sharing situations we are interested in supporting. Furthermore, these kinds of names would be very difficult for a computer system to automatically generate, and so we consider them beyond the scope of this work.

Our participants also provided some insights into the factors influencing why they named places in a certain way. A primary factor was how familiar the recipient of the information was with a given location. This observation is similar to one by Consolvo et al [9], that if people obfuscated their location, it was usually not for privacy reasons but for reasons of understandability of the recipient. Other factors mentioned by participants included simplicity of the place name (the shorter the better), and reducing potential ambiguity (“Starbucks” may cause ambiguity due to the number of Starbucks cafes, if no street or area is included).

This pilot study provided us the first taste of the preferences users have for place naming and helped us better design and organize the follow-up formal user study.

### *3.1.2 Formal Study*

We conducted another user study over the course of two weeks in March 2009. More than 80 people within our university community signed up. We selected a group of 30 users from this pool, 43.3% females and 56.7% males, with an average age of 24 years old. Among all the participants, 8 were undergraduate students, 2 university staff, 2 visiting scholars, and the rest graduate students. Our participants also worked or studied in different disciplines including science, business, engineering and arts. This was intended to maximize the diversity of participants.

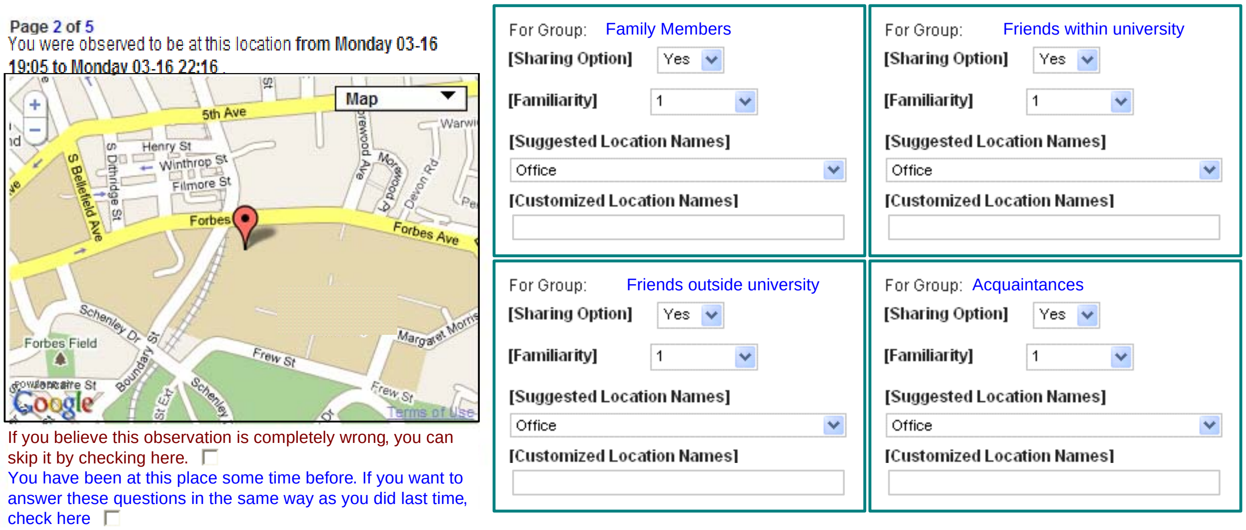


**Figure 1: A screen shot of the web application displaying a location observation between 19:05 and 22:16. This interface was presented to our study participants after uploading a location trace, to give them the chance to express how they would name a place to different people in their social network.**

We provided each participant with a Nokia N95 smart phone for one week at a time (15 participants per week, 2 weeks total, thus 30 people total). Our participants were asked to use their own SIM cards and use the N95 as their primary mobile phone for a week. We chose this approach so as to ensure that our participants carried the N95 with them and keep the phone charged. We also installed a location sensing application, which recorded the phone's location based on GPS data if satellite signals were available, or WiFi MAC addresses and signal strengths otherwise (see next section for more details).

Participants were asked to specify several groups of people in their social networks. We provided five predefined groups to them: (1) Family members, (2) Friends within university; (3) Friends outside university; (4) Superiors (including boss, advisor and so on) and (5) Other acquaintances. Participants could select from the above five groups or they could define their own social groups.

Participants were also asked to upload a file containing their location information from their phones each day. They were then asked to indicate their place naming preferences by answering several questions regarding the places they visited during that day (see Figure 1). The set of questions included (1) whether they would be willing to share that location with different groups of people, (2) how familiar the members in each group were with that location, and (3) the specific place names they would like to use. For each place, this set of questions would be repeated for different social groups, to whom the information would be conveyed.

Participants were paid a total of $30 to compensate them for their participation in the study.

## 3.2 Tools

The tools that were used in the user study included a location sensing application written for the Nokia N95 phones, and a web application that allowed participants to upload and audit their location information.

### 3.2.1 Location Sensing Application

To capture location traces, we used the LocationTracker, a location recording application previously used in Benisch's expressiveness study [6]. The application is written in C++ for Nokia's Symbian OS. It runs continuously in the background without any other user involvement. This application combines GPS and Wi-Fi positioning technologies. The phone's geo-coordinates are recorded every 15 seconds if the embedded GPS unit is able to determine its position. If the GPS signal is weak, for example if the participant is indoors or the sky is cloudy, the application records WiFi MAC addresses every 3 minutes instead. To reduce power consumption, the application uses the N95's accelerometer to trigger the positioning module. We used Skyhook API[8] to translate Wi-Fi MAC addresses into geo-coordinates when the location files were uploaded by participants.

### 3.2.2 Daily Annotation Task

Every day participants were asked to login on our web site to upload their location file and annotate the location they visited that day.

**Identify significant places.** We processed each location file by iterating through the GPS and WiFi readings to extract what we called "significant places". We defined a significant place as any location where the participant stayed for more than 5 minutes. A tradeoff of this approach is that we cannot capture names for places where people spend a brief amount of time, such as picking up packages in post office or withdrawing cash at ATM machine. However, our data suggested that our approach would still account for where a person was for the vast majority of time in a day, and these missing points would not significantly affected the results of our study, so we felt that this was a reasonable tradeoff.

**Generate place names.** For each significant place identified, we also provide some potential place names that participants could choose to save typing time and to see if simple approaches were potentially useful. We used several on-line resources to come up with this suggested list:

- Geonames[9] provides a reverse-geocoding service, translating geo-coordinates into addresses with different granularity, for example, street address, intersections, neighborhood, and so on.
- Whitepage[10] provides the name of a business given a street address. Whitepage's APIs provided a convenient way to generate place names based on function and business nature of certain locations.
- Wikipedia[11]: Many entries for places in Wikipedia are tagged with latitude and longitude. We collected the names of landmarks and points of interests in nearby cities from the Wikipedia and added them to our own database before the study.

---

[8] Skyhook Wireless, http://skyhookwireless.com
[9] Geonames, http://www.geonames,org
[10] Whitepage, http://www.whitepage.com

| Aspect | Attributes | Type |
|---|---|---|
| Participant | age<br>gender<br>status | int<br>category<br>category |
| Place | distanceFromHome<br>isOnCampus | double (km)<br>boolean |
| Time | duration<br>timeClass<br>frequency | int (sec)<br>category<br>int |
| Other | share<br>familiarity<br>socialGroup | boolean<br>int<br>category |

**Table 1 : Table of data collected in our user study. In addition to position, we also had data on the participant, the place, the time, and other factors related to the recipient of the data (would the participant have share the location, how familiar the recipient was with the location, and what social group the recipient was in relative to the participant).**

- Wikimapia[6] is a grassroots effort to name places. It has a large number of place names that are tagged to a map by its users.

We also provided several generic place names, like “Home”, “Office”, “School” and et al., which are usually used in our daily life to describe the frequently visited places.

**Annotate traces.** After a participant’s location file was processed, our web application took the participants through a series of pages which displayed the significant places she had visited in chronological order (see Figure 1). The participant’s location was shown as a marker at the center of a map with the arrival and departure time indicated above. The participant could skip this place if she believed that the location observation was inaccurate. For a repeated place where the participant had annotated before, she could choose to keep the same answers as last time without filling all the answers again. A screen shot of the user interface of the web application is shown in Figure 1.

Next to the map on each page, our web application displayed a set of questions for different groups of individuals pre-defined by the participant. For each group, the participant needed to indicate:

1) **[Sharing Option]:** Whether she would be comfortable sharing this location during the indicated time with the given group. *(yes/no)*
2) **[Familiarity]:** How familiar the members in this group were with this place? *(1-5 scale, ‘1’ means not familiar at all, ‘5’ means very familiar, ‘not sure’ if she didn’t know the answer.)*
3) **[Suggested Place Names]:** Related on-line resources were used to generate a suggested list of place names, as described above. Participants could select the place name from the generated list to save typing if they would use the exact words to refer to that location.
4) **[Customized Place Names]:** If none of the suggested place names were appropriate, the participants could assign his or her own place name to that place by typing it in the text input box.

[11] Wikipedia, http://www.wikipedia.org

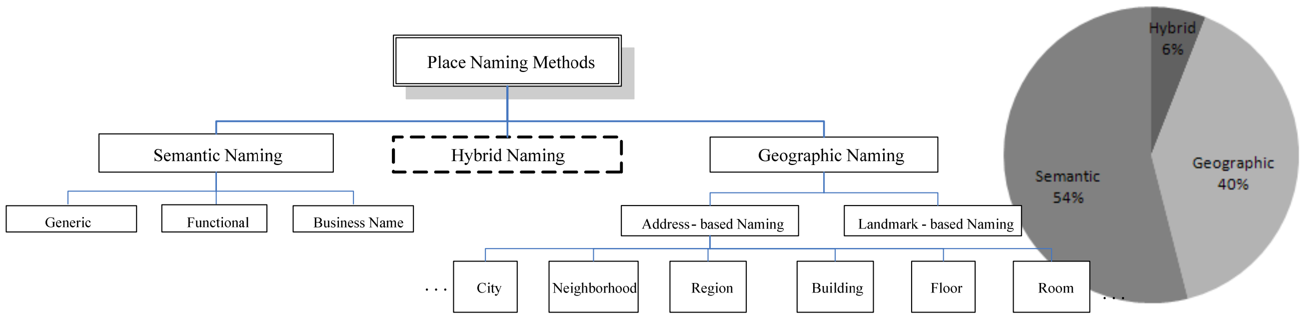


**Figure 2 : Classifications on place naming methods**

The right hand side of Figure 1 shows an example screen shot of sets of questions for four different groups.

# 4 RESULT ANALYSIS OF PLACE NAMING

In this section, we will present the results we collected and discuss the implications behind these numbers.

## 4.1 Descriptive Statistics

LocationTracker captured 66014 location readings from our 30 participants. From these location readings, we extracted 408 unique places where participants spent more than 5 minutes. On average, each participant visited 13.53 distinct places over a week, with a standard deviation of 4.98. It is worth mentioning that three participants travelled outside the city where the study took place during the one-week period. Besides the place names given to each extracted place, eleven different attributes were also captured, which covered several major aspects of location sharing (see Table 1). Again, our goal was to identify significant dimensions among all these captured attributes which greatly influence the way people name a place.

## 4.2 Taxonomy of Place Naming

By analyzing all the place names collected in our study, we were able to identify several general place naming patterns based on different kinds of information implied. We further categorized these place naming methods into a hierarchical classification shown in Figure 2 and examples of each category listed in Table 2. First, we classify the place naming methods into three top-level categories: *semantic naming*, *geographic naming*, and *hybrid naming*.

*Semantic naming* provides place names with certain semantic meanings, for example, ‘home,’ ‘coffee shop,’ and ‘Barnes & Noble.’ Places identified by semantic naming methods tend to be difficult pinpoint (or uniquely pinpoint) on a map without extra knowledge. Under this top-level category, we observed three widely used sub-categories. The first sub-category is *generic naming*, which refers to terms commonly used within almost all contexts, such as ‘home’ and ‘work’. The second sub-category is *functional naming*. Place names under this category reveal the functional nature of those places or imply the activities carried out at those spots. Examples include ‘restaurant,’ ‘gym,’ and ’church.’ The last sub-category is *business name*, which is

| Categories | Examples |
|---|---|
| Semantic | |
| Generic | Home, work, school |
| Functional | Restaurant, gym, lab |
| Business name | Starbucks, Barnes & Noble, Rite Aid |
| Geographic | |
| Address-based | New York city, Shadyside neighborhood, Waterfront, CIC building 231 |
| Landmark-based | Near Liberty Bridge, Next to Frick Park |
| Hybrid | |
| | Giant eagle on Murray Ave, USPS in Squirrel Hill |

**Table 2: Examples of each place naming method. We identified three top-level ways for how people named places (semantic, geographic, and hybrid). We also identified several sub-categories.**

similar to *functional name* but directly quotes the registered business name, such as ‘Barnes & Noble’ and ‘Starbucks’. The function of these places can be implied if people have knowledge on the registered name. Business name and functional names could also be used together, like ‘Rite Aid grocery store’.

In contrast, *geographic naming*, the second top-level category, refers to a certain area defined on map. Two sub-categories here are *address-based naming* and *landmark-based naming*. Address naming uses the location’s address or part of the address as the place name. On the other hand, *landmark naming* uses a nearby well-known spot or other public places to refer to the target location, like ‘near Liberty Bridge’ or ‘next to Central Park’.

The third top-level category is *hybrid naming*, which combines the features of both semantic naming and geographic naming. Examples include ‘Starbucks on Craig St,’ ‘Barnes & Noble in Squirrel Hill.’ Hybrid naming is usually used to eliminate the ambiguity caused by using semantic naming alone.

The pie chart in Figure 2 indicates the breakdown for each of the three top-level categories for all the place names used by our participants. Note that for a single place, the naming methods can change in different situations. For example, a participant could call a certain place “Starbucks” but refer to it as “coffee shop” later. Similarly, the naming methods were also likely to be different when the place was referred to different groups of people. Based on this classification scheme, we labeled all the annotation records with the place naming methods and granularity levels.

## 4.3 Predicting Place Naming Method

We collected 3376 place naming records from our participants for all the 408 places. Each record includes a vector $x$ that corresponds to 11 attributes listed in Table 1 and a label $y$ which represents one of the five sub-categories introduced in 4.2 (generic, functional, business name, address-based, landmark-based place naming)[12]. For all the records with label 'address-based naming,' another label *y'* is attached to indicate its granularity. We used an integer number within the range of [0, 8] to represent the granularity, with 0 representing country level granularity (coarsest) and 8 representing room level granularity (finest). In our study, we observed the granularity used by our participants ranging from 3 (city level granularity) to 8 (room level granularity). Both the *y* and *y'* were manually labeled by our researchers according to the classification scheme.

People's preferences for place naming are complex and dynamic. They are difficult to be accurately captured by single attribute or a couple of attributes, thus it is impossible to create a single decision tree to predict the place naming method used in different situations. Random forest algorithm provides more flexibility in classification by considering multiple possible decision trees [7]. To classify a new object from an input vector, this algorithm puts the input vector down each of the trees in the forest. Each tree gives a classification, and the tree "votes" for that class. The forest chooses the classification having the most votes over all the trees in the forest. We used Weka[13] tool kit to train and test the random forest classifiers on both the whole dataset and individual dataset (data associated with each individual participant). Datasets were randomly spited into training set and testing set with the ratio 2:1, and classification results were averaged over 10 runs.

The classifier trained over all participants gave an average classification accuracy of 93.7%. We also trained specific classifier for individual participant. The classification accuracies of these classifiers are shown in Figure 3. The average classification accuracy is 91.6%. We also used the random forest to predict place naming granularity for the records that were annotated by using address-based naming for all the participants and achieve an average accuracy of 84.5%. (The dataset was not large enough to enable training classifiers for individual participant.)

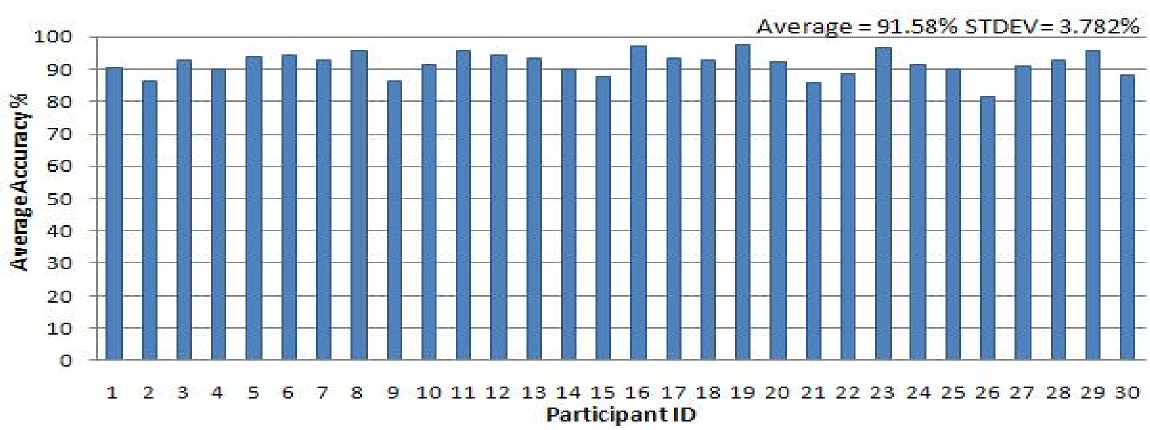


**Figure 3 : Classification accuracy of place naming methods for each participant**

[12] Records contain hybrid place names were removed in order to make sure each record associated with exactly one label, remaining 3173 records in the dataset.

We concluded that people's preferences for place naming are complex and dynamic but also predictable in terms of naming methods and naming granularity. We believe that by accurately predicted these two factors automatically generating place names would become much easier.

## 4.4 Analysis of Patterns in Place Naming

While machine learning techniques provide a black box to predict the place naming method and granularity, we are more interested in understanding the relationship between the attributes we collected and the way people name a place. To this end, we used statistical methods to analyze the data.

### *4.4.1 Statistical Mode*

Since most regression models assume that attributes are independent of each other, we used Weka [13] to do Pearson's chi-square test on each pair of attributes. The test results suggested that there was no strong correlation between attributes; thus we could assume all the attributes were independent. The independence of attributes also guarantees that the influence of each attribute is isolated, hence makes the interpretation of regression results more accurate. We used logistic regression to model the relationship between the binomial response variable ($y$) and all the explanatory variables ($x$) that were either continuous or categorical. All the numerical attributes were normalized before being inputted into the model. Each of the regression estimates (regression coefficients) describes the size of the contribution of corresponding attribute. A positive estimate means that the attribute increases the probability of the outcome, while a negative estimate means that the attributes decreases the probability of that outcome.

Since multiple records were generated by one participant and thus likely to be correlated, we chose to use a *Logistic Regression for Repeated Measures* (LRRM) [3] in our analysis. This model adopts the method of generalized estimating equations (GEE) [12] that can account for correlations among records from the same participant. In our analysis, p-values were used to select significant attributes. Attributes are more significant if the corresponding p-values are smaller. We adopted the standard threshold of *p-value=0.05* to filter out unimportant attributes. All attributes with p-value greater than 0.05 were dropped.

[13] Weka, http://www.cs.waikato.ac.nz/~ml/weka/

### *4.4.2 Interpretation of Regression Results*

**Semantic Naming vs. Geographic Naming**

Since the hybrid naming method only takes 6% of all place names, we eliminated these records and focused on distinguishing between the semantic naming (S) and geographic meaning (G). So the response variables y consists of two possible categorical values, S and G. Explanatory variable x included all the 11 attributes at the very beginning. Insignificant attributes would be dropped latter based on their p-values.

Table 3 presents the results of logistic regression for repeated measures. The first column shows the names of important attributes. The second column lists the regression estimate of corresponding attributes. For each attribute, if the estimate is positive, the probability of using a semantic name *Pr(S)* increases when the value of this attribute increase (in other words, Pr(G) decreases); if the estimate is negative, the probability of using a semantic name *Pr(S)* decreases when the value of this attribute decreases. The p-values listed in the third column represent the significance of the corresponding attributes. We only list the attributes with p-value smaller than 0.05.

Table 3 suggests that when other attributes stay unchanged, the *Pr(S)* increases if (1) the distanceFromHome decreases; (2) the frequency increases; (3) the value of attribute isOnCampus becomes true; (4) the timeClass attribute changes from 'weekend' or 'after work' to 'working hour'; (5) the share attribute becomes false; (6) the value of socialGroup changes to categories associated

| Attribute | Estimate | p-value |
|---|---|---|
| distanceFromHome | -11.117 | **<.0001** |
| frequency | 8.4183 | **0.0044** |
| isOnCampus | | |
| 0:false | 0.0000 | **<.0001** |
| 1:true | 0.7605 | **.** |
| timeClass | | |
| 0: working hr | 0.0000 | **<.0001** |
| 1: after work | -0.0590 | **<.0001** |
| 2: weekend | -0.7172 | **<.0001** |
| share | | |
| 0:false | 0.0000 | **<.0001** |
| 1:true | -1.2699 | **<.0001** |
| socialGroup | | |
| 0: Family | 0.0000 | **<.0001** |
| 1: Friends within university | -0.1678 | **0.0235** |
| 2: Friends outside university | -0.6130 | **<.0001** |
| 3: Superiors | -0.4492 | **0.0331** |
| 4: Acquaintances | -0.7865 | **<.0001** |

**Table 3 : LRRM results on probability of using semantic naming method -- *Pr(S)*. For each attribute, the sign of each regression estimate represents the positive or negative influence of that attribute. E.g. the larger distanceFromHome value, the smaller *Pr(S)*.**

| Attribute | Estimate | p-value |
|---|---|---|
| familiarity | 12.4727 | **<.0008** |
| frequency | 10.400 | **0.0001** |
| isOnCampus | | |
| 0:false | 0.0000 | **<.0001** |
| 1:true | 2.3120 | **<.0001** |
| socialGroup | | |
| 0: Family | 0.0000 | **0.0193** |
| 1: Friends within university | -0.6006 | 0.1530 |
| 2: Friends outside university | -0.6786 | 0.5562 |
| 3: Superiors | -1.1176 | 0.2637 |
| 4: Acquaintances | -1.5448 | 0.0654 |

**Table 4: LRRM results on probability of using business names – *Pr(B)*.**

| Attribute | Estimate | p-value |
|---|---|---|
| distanceFromHome | -3.096 | **<.0001** |
| frequency | 0.1857 | **<.0001** |
| share | | |
| 0:false | 0.0000 | **0.0065** |
| 1:true | 0.1460 | **<.0001** |
| socialGroup | | |
| 0: Family | 0.0000 | 0.2162 |
| 1: friends within university | -0.0061 | 0.0653 |
| 2: friends outside university | -0.5493 | **0.0073** |
| 3: Superiors | -0.0075 | 0.7835 |
| 4: Acquaintances | -0.0087 | 0.2356 |

**Table 5 : Linear regression results on granularity.**

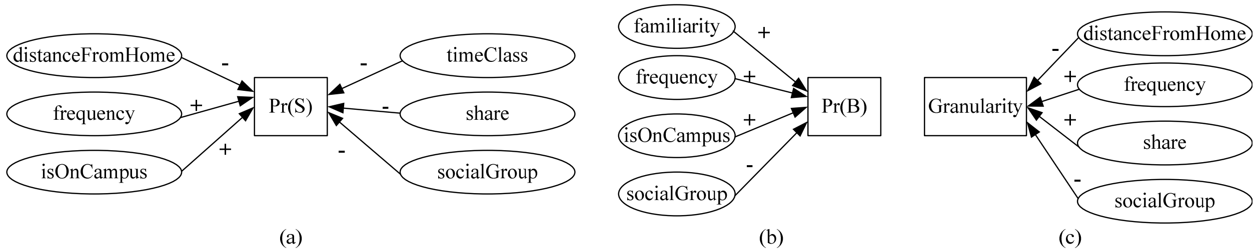


**Figure 4 : Influence graph extracted from Table 3-5: (a) influence of 6 attributes on the probability of using semantic naming method based Table 3; (b) influence of 4 attributes on the probability of using business name based on Table 4; (c) influence of 4 attributes on the granularity based on Table 5. In each sub-figure, '+' sign indicates positive impact of the corresponding attribute and '-' sign indicates negative impact of the corresponding attribute.**

with smaller numbers. These influences are summarized in Figure 4 (a), where '+' and '-' signs indicated the positive (+) or negative (-) impact of each attribute. Figure 4(a) illustrated that frequency, isOnCampus attributes have positive influence on *Pr(S),* while the other four attributes have negative influence on *Pr(S).*

We can conclude that the semantic naming methods are more likely to be used in the following situations:

- When place names are used to refer to routine places (frequently visited , close to home or work place, in work days)
- When people are not willing to share their location with others
- When recipients of this location information have a closer relationship with the one who shares this information

### Generic Naming is Relatively Context Independent

We also found that, among the three sub-categories under semantic naming, the generic naming method is relatively context independent. For example, the terms like "home" and "office" were consistently used by our participants in almost all situations. We identified P as a place set which included all the places that had one or more annotation records labeled by the generic naming method. We then calculated the ratio of the number of annotation records with label 'generic naming' over all the number of records associated with places in the set *P*. An average percentage of 94.93% with 3.25% standard deviation was observed. In other words, the places in set P were named by generic naming method consistently for 94.93% of time on average. Therefore we can claim that the generic naming method is context independent with all non-place-related attributes.

### Business Naming vs. Functional Naming

We also applied LRRM to understand what attributes influenced the decision of using a business name (B) or a functional name (F). Table 4 suggests that when other attributes stay unchanged, the *Pr(B)* increases if (1) the familiarity value increases; (2) the frequency increases; (3) the value of isOnCampus changes from false to true; (4) the value of socialGroup changes to categories associated with smaller numbers. These influences are summarized in Figure 4 (b). We conclude that business name is more likely to be used in the following situations:

- When the place names are used to refer to more frequently visited places.

- When the recipients of this location information share more common knowledge (closer relationship with the sender, more familiar with this place.)

**Granularity**

For all the records annotated by address-based naming, we assigned a granularity level *y'* associated with each place name. The smaller the *y'* value, the coarser the granularity. Linear regression was used here since the *y'* value has numerical meaning. The regression results and interpretation of each attributes are illustrated in Table 5. When other attributes stay unchanged, the granularity increases if (1) the value of distanceFromHome decreases; (2) the frequency increases; (3) the value of share changes from false to true; (4) the value of socialGroup changes to categories associated with smaller numbers. These influences are summarized in Figure 4 (c). To recap,

- When people are far away from their home area (e.g. travel outside the city they live), coarser granularity are more often used to refer to places.
- When people have more concerns on their privacy, they tend to use place names with coarser granularity.

# 5 DISCUSSION

## 5.1 Presentation of Location Information Influence People's Privacy Concern

In Tsai's study [33], the authors showed the statistical results of people's openness of sharing their location with social groups on Facebook[14]. Their results indicated that participants were willing to share their location 72.9% of time on average. In Benisch's expressiveness study [6], he also reported the openness of sharing location information in terms of the percentage of time they were willing to disclose their location to different social groups. They found that on average subjects would be comfortable sharing their locations about 89% of the time with friends, 86% of the time with family, 46% of the time with other individuals in their university and 26% of time with the general population. In our study, participants were willing to share their location 89.8% of time with others on average. Since our study and Benisch's study had different predefined social groups, it may not be fair to conduct group-wise comparisons. However, since the openness (89.8%) of our participant across all groups has already exceeded the highest openness (89%) among four social groups in their work, we consider that our participants have more open attitudes towards location sharing.

One crucial difference between our study and the other two studies is the presentation of location information. In their study, participants' location information was presented to recipients by displaying geo-coordinate estimates on maps. Under this setting, participants had only binary controls (share or not share) in sharing their location information. In contrast, we used place names to present the location information. People have more choices of how to modulate the information they would like to convey to others, thus their privacy preferences were better accommodated. Therefore, people tended to be more comfortable when sharing locations by using place names. This finding also confirmed the privacy preserving benefit of place names we mentioned in Section 1.

[14] Facebook, http://www.facebook.com

## 5.2 Two High Level Factors Influence Place Naming

Based on the data analysis results presented in Section 4.4, we can further extract two crucial factors that influence people's place naming preferences. They are people's routine and their privacy concerns.

We refer routine places as the locations where people visit repeatedly. These places usually have relatively short distance from homes or work places. People tend to use more semantic place names when they refer to their routine places. This is because that most people share a lot of common knowledge on these places with others who have closer relationship, which guarantees the place names were understandable for the recipients.

At the same time, people's willingness of disclosing location information influences the granularity of the location presentation. When people feel less comfortable to share their locations, such as sharing location information with strangers, they tend to disclose less information to the recipient. By manipulating the granularity of place names, people could adjust the amount of information they want to share with others.

Therefore, if a person's profile which includes her location history as well as privacy preferences is available, we can somehow predict the way she name a place in the context of location sharing without asking her to input location names in order to train the machine learning model.

## 5.3 Possible Bias in the Empirical Study

All the participants in our study were recruited within the university community. We made our best effort on diversifying the sample pool by selecting people from different disciplines. Although we didn't observe strong influence from participant related attributes, like age, gender or status, a user study of more participants and more diversify composition of participants would be definitely desired in order to verify the findings in our work. Moreover, in our study, participants' location information was not actually shared, but people may not behave in the same way once the chance for a release really arises.

We also realize that there might be several other attributes which were not captured in our study but may influence the way people name a place, such as the distance between the person who discloses the location information and the person who views the location information, the specific purpose of sharing location information. Also, we need to consider some abnormal situations, in which people's place naming preferences dramatically changes for a short period of time. For example, being late for a meeting can lead to a disclosure of location information with very fine-grained granularity, which is not likely to happen in normal situation. These attributes and situations are difficult to capture in a user study that lasts for several days. A long time deployment of a real location-sharing system that features place name presentation would be a better way to study these factors.

## 6 CONCLUSIONS AND FUTURE WORK

Most existing location sharing applications present users' location information by showing it on a map. On the contrary, sharing location information in the form of place names can provide more meanings and preserve users' location privacy better. However, the dynamic feature of place name makes automatically generating place names a research challenge.

In our work, we studied the people's preferences for place naming through a week-long user study with 30 participants. Based on all the place names we collected, we proposed a hierarchical classification of place naming methods according to different information implied by place names. By applying machine learning techniques, we were able to predict the way people used to refer to certain place with an average accuracy of 93.7% and predict the granularity of place names with an accuracy of 84.5%. We further identified that people's routine and privacy concerns are the two crucial factors which strongly influence the way people name a place. We also observed that people tend to be more willing to share their location in the form of place names than to share their exact positions.

Our findings provide important practical implications to the research problem of how to automatically generate place names. We will explore additional dimensions including some dynamic factors that might influence place naming in a larger scale empirical study with more diversified participants. Future work could also be conducted on designing, building, deploying and evaluating a location sharing system which presents location information in the form of dynamically generated place names.

## 7 ACKNOWLEDGMENTS

This work is supported by NSF Cyber Trust grant CNS-0627513 and ARO research grant DAAD19-02-1-0389 to Carnegie Mellon University's CyLab. Additional support has been provided by Microsoft through the Carnegie Mellon Center for Computational Thinking, FCT through the CMU/Portugal Information and Communication Technologies Institute, and through grants from FranceTelecom and Nokia. The authors would also like to thank Erick Fang, Michael Benisch, Patrick Gage Kelley, Janice Tsai, Eran Toch, and Jackie Xiao for their assistance with our study.

# 8 REFERENCES


[1] D. Ashbrook, “Learning Significant Locations and Predicting User Movement with GPS,” in Proceedings of the 6th IEEE International Symposium on Wearable Computers, 2002.

[2] D. Ashbrook, and T. Starner, “Using GPS to learn significant locations and predict movement across multiple users,” *Personal Ubiquitous Comput.*, vol. 7, no. 5, pp. 275-286, 2003.

[3] A. Azzalini, “Logistic regression for autocorrelated data with application to repeated measures,” *Biometrika*, vol. 81, no. 4, pp. 767-775, 1994.

[4] L. Barkhuus, B. Brown, M. Bell *et al.*, “From Awareness to Repartee: Sharing Location within Social Groups,” in CHI, 2008, pp. 497-506.

[5] L. Barkhuus, and A. Dey, “Location-based Services for Mobile Telephony: a Study of Users' Privacy Concerns,” in INTERACT, 2003, pp. 702-712.

[6] M. Benisch, P. G. Kelley, N. Sadeh *et al.*, *The Impact of Expressiveness on the Effectiveness of Privacy Mechanisms for Location Sharing,* CMU-ISR-08-141, Carnegie Mellon University, 2008.

[7] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[8] S. Consolvo, I. E. Smith, T. Matthews *et al.*, “Location disclosure to social relations: why, when, & what people want to share,” in Proceedings of the SIGCHI conference on Human factors in computing systems, Portland, Oregon, USA, 2005.

[9] J. Cornwell, I. Fette, G. Hsieh *et al.*, “User-controllable Security and Privacy for Pervasive Computing,” in 8th IEEE Workshop on Mobile Computing Systems and Applications, 2007.

[10] P. Fagerberg, F. Espinoza, and P. Persson, “What is a place?: allowing users to name and define places,” in CHI '03 extended abstracts on Human factors in computing systems, Ft. Lauderdale, Florida, USA, 2003.

[11] R. Genereux, L. Ward, and J. Russel, “The Behavioral Component in the Meaning of Places,” *Journal of Environmental Psychology*, vol. 3, pp. 43-55, 1983.

[12] J. W. Hardin, and J. Hilbe, *Generalized Estimating Equations*: CRC Press, 2002.

[13] J. Hightower, “From Position to Place,” in The 2003 Workshop on Location-Aware Computing, 2003.

[14] J. Hightower, S. Consolvo, A. Lamarca *et al.*, "Learning and Recognizing the Places We Go," *UbiComp 2005: Ubiquitous Computing*, pp. 159-176, 2005.

[15] J. I. Hong, and J. A. Landay, “An architecture for privacy-sensitive ubiquitous computing,” in Proceedings of the 2nd international conference on Mobile systems, applications, and services, Boston, MA, USA, 2004.

[16] S. Huang, F. Proulx, and C. Ratti, “iFIND: a Peer-to-Peer Application for Real-time Location Monitoring on the MIT Campus,” in CUPUM, 2007.

[17] G. Iachello, I. Smith, S. Consolvo *et al.*, “Control, Deception, and Communication: Evaluating the Deployment of a Location-enhanced Messaging Service,” in UbiComp, 2005, pp. 213-231.

[18] C. Jiang, and P. Steenkiste, “A Hybrid Location Model with a Computable Location Identifier for Ubiquitous Computing,” in Ubiquitous Computing '02, 2002.

[19] B. Kramer, “Classification of Generic Places: Explorations with Implications for Evaluation,” *Environemental Psychology,* vol. 15, pp. 3-22, 1995.

[20] J. Krumm, and E. Horvitz, "Predestination: Inferring Destinations from Partial Trajectories," *UbiComp 2006: Ubiquitous Computing*, pp. 243-260, 2006.

[21] E. Laurier, “Why People Say Where They are During Mobile Phone Calls,” *Environment and Planning D: Society and Space*, 2000.

[22] S. Lederer, J. Mankoff, and A. K. Dey, “Who Want to Know What When? Privacy Preference Determinants in Ubiquitous Computing,” in CHI, 2003, pp. 724-725.

[23] L. Liao, *Location-based Activity Recognition,* 1269296, University of Washington, 2006.

[24] L. Liao, D. Fox, and H. Kautz, “Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields,” *International Journal of Robotics Research,* vol. 26, no. 1, pp. 119-134, 2007.

[25] J. Liu, O. Wolfson, and H. Yin, “Extracting Semantic Location from Outdoor Positioning Systems,” in Proceedings of the 7th International Conference on Mobile Data Management, 2006.

[26] E. Miluzzo, N. Lane, S. Eisenman *et al.*, "CenceMe: Injecting Sensing Presence into Social Networking Applications," *Smart Sensing and Context*, pp. 1-28, 2007.

[27] L. Mummidi, and J. Krumm, “Discovering points of interest from users’ map annotations,” *GeoJournal*, vol. 72, no. 3, pp. 215-227, 2008.

[28] T. Rattenbury, N. Good, and M. Naaman, “Towards automatic extraction of event and place semantics from flickr tags,” in Proceedings of the 30th annual international ACM SIGIR, Amsterdam, The Netherlands, 2007.

[29] N. Sadeh, F. Gandon, and O. B. Kwon, *Ambient Intelligence: The myCampus Experience.* , Technical Report CMU-ISRI-05-123, Carnege Mellon University, 2005.

[30] N. Sadeh, J. Hong, L. Cranor *et al.*, “Understanding and Capturing People's Privacy Policies in a People Finder Application,” *The Journal of Personal and Ubiquitous Computing*, 2008.

[31] E. Schegloff, “Notes on a Conversational Practice: Formulating Place,” *Studies in Social Interaction*, 1971.

[32] B. N. Schilit, and M. M. Theimer, “Disseminating active map information to mobile hosts,” *Network, IEEE,* vol. 8, no. 5, pp. 22-32, 1994.

[33] J. Y. Tsai, P. Kelley, P. Drielsma *et al.*, “Who's Viewed You? The Impact of Feedback in a Mobile Location Sharing System,” in CHI, 2009.

[34] J. Wang, and J. Canny, “End-user place annotation on mobile devices: a comparative study,” in CHI '06 extended abstracts on Human factors in computing systems, 2006.

[35] Y. Wang, J. Lin, M. Annavaram *et al.*, “A Framework of Energy Efficient Mobile Sensing for Automatic User State Recognition,” in MobiSys, 2009.

[36] R. Want, A. Hopper, and J. Gibbons, “The active badge location system,” *ACM Trans. Inf. Syst.*, vol. 10, no. 1, pp. 91-102, 1992.

[37] A. Weilenmann, “"I can't talk now, I'm in a fitting room": Formulating Availability and Location in Mobile Phone Conversations,” *Environment and Planning*, vol. 35, no. 9, pp. 1589-1605, September, 2003.

[38] M. Yutaka, O. Naoaki, I. Kiyoshi *et al.*, “Inferring Long-term User Property based on Users' Location History,” in IJCAI, 2007.

[39] C. Zhou, D. Frankowski, P. Ludford *et al.*, “Discovering personally meaningful places: An interactive clustering approach,” *ACM Trans. Inf. Syst.*, vol. 25, no. 3, pp. 12, 2007.

[40] C. Zhou, P. Ludford, D. Frankowski *et al.*, “An experiment in discovering personally meaningful places from location data,” in CHI '05 extended abstracts on Human factors in computing systems, Portland, OR, USA, 2005.

[41] C. Zhou, P. Ludford, D. Frankowski *et al.*, "How Do People’s Concepts of Place Relate to Physical Locations?," *Human-Computer Interaction - INTERACT 2005*, pp. 886-898, 2005.

[42] C. Zhou, P. Ludford, D. Frankowski *et al.*, "Talking about Place: An Experiment in How People Describe Places."