Understanding People's Place Naming Preferences in Location Sharing

Jialiu Lin, Guang Xiang,
Jason Hong, Norman Sadeh

March 2010
CMU-ISR-09-121R
CMU-CyLab-09-010R

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This is an updated version of CMU-ISR-09-121/CMU-CyLab-09-010.

ABSTRACT

Most location sharing applications display people's locations on a map. However, in practice, people use a rich variety of terms to refer to their locations when interacting with others, such as "home," "Starbucks," or "the bus stop near my house." Our long-term goal is to create a system that can automatically generate appropriate place names based on real-time context and user preferences. As a first step, we analyze data from a two-week study involving 26 participants in two different cities, focusing on how people refer to places in location sharing. We derive a taxonomy of different place naming methods, and show that factors such as a person's perceived familiarity with a place and the entropy of that place (i.e. the variety of people who visit it) strongly influence the way people refer to it when interacting with that person. We proceed with the description of a machine learning model for predicting how people name places. Using our data, this model is able to predict the place naming method people choose with an average accuracy higher than 85%.

1 INTRODUCTION

The past few years have seen the launch of a growing number of "friend finder" applications which let people share their location with others [2, 4-6, 14, 23, 38, 46]. Many of these applications 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, they fail to capture the nuances people often use when referring to their location in interactions with others. People usually do not describe their locations to others as, for example, "40.443 north, -79.941 west" or "5837 Centre Ave." Instead, they often rely on a wider range of terms and nuances such as "home," "Starbucks," "near Liberty Bridge," or "Chicago." These kinds of place descriptions allow people to modulate the amount of information they disclose and adjust the way they describe their location to account for both privacy and utility considerations – the latter referring to how useful a given piece of information is likely to be to a particular individual in a given context. These examples illustrate the complex nature of place naming. A given location may be referred to in different ways depending on the situation.

Being able to computationally generate place names that capture these nuances could make location sharing applications more useful, enabling people to share more meaningful information with one another based on particular circumstances and giving them a wider range of privacy options when it comes to controlling how much information they disclose. For example, a person might be willing to let some people know they are at "home", but uncomfortable showing them their home on a map or disclosing its street address. In addition, generating meaningful place names could render the integration of location information with other services more valuable. For example, a person could share her current location as a status message in an instant messaging client or on a social networking site, or show a text label denoting the place a photo was taken in a photo sharing application. This level of integration is less meaningful and/or desirable when location information is limited to what generally amounts to a dot on a map.

In short, today there is a gap between how people actually name places and what technology can offer [51]. Reverse-geocoding systems can translate geo-coordinates into street addresses, neighborhoods, postal codes, and countries, but these kinds of names only provide information from a geographical perspective, hence do not always match the way people would want to refer to places. As a first step towards building a place naming system, we collected data through a two-week study with 26 participants in two different cities (Pittsburgh & Moffett Field), where we examined preferences for how people name places. We recorded the location traces of our participants over this time period, and followed up with participants to understand what factors influenced how they named the places they visited. By analyzing and modeling all the place names collected in our study, we were able to identify several general patterns. In brief, this report makes the following research contributions:

- By positioning place naming into a hierarchical framework, we identify two major methods that people use to tailor the place names they want to disclose in location sharing, namely choosing a *perspective to describe the place* (semantic, geographic, or hybrid) and tuning the *granularity of disclosure*.
- We identify factors that influence the way people refer to a location, including some factors that had not been examined previously, such as a *recipient's perceived familiarity with the location* (in the sharer's view) and a *location's entropy*, a measure that estimates how many different people visit that place.
- By applying machine learning to model people's place naming preferences, our approach offers more flexibility and effectiveness in predicting the method and granularity of how people refer to a place, with an average accuracy higher than 85% in our experiments.

2   RELATED WORK

To the best of our knowledge, little work has been done in generating place descriptions according to different contexts or in statistically modeling people's preferences. However, there are several directions closely related to place naming. We have organized the work into five themes: contextual meaning of locations, location sharing applications, place discovery, computing models of places, and grassroots place labeling.

2.1   Contextual Meanings of Places

In the 1970s, researchers in social interaction and environmental psychology documented several underlying meanings of locations [30, 40, 47]. A meaningful place name can capture the location's demographic, historic, environmental, personal, as well as commercial significance [20]. When supplemented with other knowledge, location information can also be used to infer higher level contextual information, such as a person's activity, level of availability or interruptibility (see, for example, [19, 28, 35, 45, 48]).

An important observation regarding place descriptions is that a person can associate multiple place names to the same place, depending on the situation and the kind of information that person wants to address. Zhou et al.[52] pointed out this dynamic feature of place descriptions and investigated the types of descriptions people naturally produce for places in a qualitative manner. However, they only reported these observations without further analysis or modeling on the collected data. In Connecto [11], Barkhuus et al. pointed out four different types of location labels participants used in the study, i.e. (1) geographic references, (2) personal meaningful place, (3) activity-related labels, and (4) hybrid labels. Their classification provides us great insights in how to classify place names. We further augment this classification by adding more fine-grained categories and organizing them into a hierarchy.

The key difference with our work in this report from this past work is that we are focused on quantitatively understanding how people name places to different people in different situations, and building a machine learning model that can support this process.

2.2   Location Sharing Applications

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, 6, 7, 11, 12, 14, 15, 23, 38, 43, 46]. Researchers found that people have significant privacy concerns when sharing their location with others [11, 12, 16, 22, 24, 31, 39]. Iachello et al. argued that it is essential for applications to support plausible deniability when disclosing location information. They also designed and evaluated Reno [25], a location-enhanced mobile coordination tool and person finder. In Reno, users were allowed to define their own names for places (such as "home" or "office") and associate them with a specific location. However, this process was not automated, thus requiring a lot of user involvement.

Some recent applications provide users more control of their privacy preferences [39, 42], such as the application mentioned by Cornwell et al. [16], the later version of which is also called "Locaccino" [5]. Locaccino is a user-controllable location sharing tool which gives users 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. These improved friend finder applications give users controls on when, where, to whom their location should be disclosed, but seldom do they provide mechanisms on how the location information is presented, hence users have limited control on tailoring the information they want to emphasize.

The Whereabouts clock developed by Brown et al. [14] shared coarse-grained semantic location among family members. Their study demonstrated the usefulness of location sharing in improving family life. Their study also suggests a strong motivation for sharing generic place names. However, it is not clear whether their findings can be generalized to social groups other than family members.

The work by Consolvo et al. [15] is the most relevant one to our report. They designed a series of ESM studies to explore whether users were willing to share their location with others, as well as what they would share. They argued that the information disclosed depended primarily on the relationship between the sharer and recipient, the purpose of sharing, and the necessary level of detail needed by the recipient. The authors also argued that utility was the primary reason for users to modulate the information. Our work builds on this past work in many ways. We exploit more attributes that haven't been covered in their study. We analyze people's place naming method in a more quantitative way with all conclusions backed up by statistical techniques. We also introduce machine learning techniques in model the data, aiming at accurately predicting people's place naming methods. Finally, we provide some evidence suggesting that privacy actually does influence what is shared, but in a subtle way.

In summary, the key difference with our work from past work is that we are not only interested in understanding users' location sharing preferences, but also in building a statistical model for automatically generating appropriate place names in different contexts.

2.3 Place Discovery

Place discovery algorithms are one way to bridge the gap between geo-coordinates and places [18, 27, 50]. Extracting significant places is also an ongoing theme in the machine learning and data mining communities [9, 10, 32-34].

For example, Ashbrook et al. extracted significant places by clustering GPS data taken over periods of time at different granularities [9, 10]. Similarly, Liao et al. successfully extracted people's activities and significant places from traces of GPS data [32, 33]. Along similar lines, Zhou et al. [49, 50] built a place discovery system based on users' location data and evaluated their system by comparing the discovery results with ground truth captured in retrospective user interviews. Hightower et al. [21, 27] used WiFi, GSM radio fingerprints as well as RF-Beacons to automatically learn the places by identifying the arrival and departure of users. Krumm et al. [29] used 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 has made good progress on clustering traces and discovering salient places, though this past work does not offer a way to automatically assign names to these recognized places. In contrast, our work is focused on paving the way towards associating meaningful names and other information with these extracted places. Our work in this report focuses specifically on modeling the data from 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.

2.4 Computing Models for Places

Schilit et al. [41] proposed a hierarchical location model to index different locations within a certain region and at different granularities, such as regions, buildings, and floors. Similarly, Jiang et al [26] proposed a computable location identifier that used a URL-like string to define the hierarchical structure of different locations.

These kind of top-down methods work well in representing a location's geographic properties. However, these methods cannot capture other semantic properties, such as what is the place's function, whether it is a public place or not. Furthermore, these kinds of top-down methods are difficult to scale up due to the tremendous effort needed to define the hierarchical structure in the first place.

2.5 Grassroots Place Labeling

An alternative way to obtain place names is by aggregating place names from grassroots contributors [20, 36]. Some location sharing applications let users give names to places, such as Reno [25] and Connecto [11]. Other location sharing application, such as "foursquare" [1] and "Gowalla"[3], adopt a "check-in" method, which basically asks users to submit the location they want to share by entering the name of the place. Check-ins require users to proactively enter the information they want to share instead of automating (or semi-automating) the process.

Websites like Wikimapia and Flickr encourage users to tag their resources, which can help in generating labels for places. For example, Rattenbury et al. [37] proposed an approach for extracting place descriptions from tags on Flickr. However, these methods also face several problems such as how to eliminate "bad" labels, how to create incentives for users to contribute, and how to preserve contributors' privacy. Wang et al. [44] proposed four different prototypes of place annotation system on mobile phones and compared their usability through a series of user studies. Their findings suggested implications on how to make a place annotation system more useful.

Grassroot labeling may be a way to gather candidate place descriptions with relatively low cost. However, this approach only partially addresses the fundamental problem we are examining in this report. More specifically, grass root labeling can provide us with a pool of potentially useful place names, but does not tell us how to select appropriate ones based on real time situations.


3   AN EMPIRICAL STUDY OF PLACE NAMING

To gather data on how people named places under different circumstances, we conducted a two-week user study in August 2009 with participants in two cities, i.e. Pittsburgh, PA and Moffett Field, CA. We collected location traces from participants and asked them what information they would like to share about their locations, based on various factors such as who was asking, how familiar the recipient was with the location, and so on. These factors are described in greater detail below.

We considered using Experience Sampling Method (ESM) to gather data, but opted for location traces for greater coverage of the places a person visited. A weakness here is that our participants had to add names to these places retrospectively, but we felt that this was an acceptable tradeoff. In addition, we felt that ESM would place a heavy burden on participants since typing on mobile devices is slow, and could negatively impact our results.

We asked participants to complete both an entrance and exit survey. The entrance survey asked participants to list the names of several people in three different social groups: *family members, close friends,* and *acquaintances.* We asked each participant to indicate the physical distance between herself and others in her social network at four different levels, i.e. *in the same city, in same state but different cities, in the same country but different states, in different countries.* Previous work [15] found that this attribute influences user's sharing behaviors. The exit survey probed participants' attitudes toward sharing location information in different forms (i.e. showing on map vs. place names). We later used the exit survey results part of the user profile to guide the data modeling.

We asked our participants to use one of our Nokia N95 smartphones as their primary cell phones (i.e. using their own SIM cards), with a location sensing application installed. We used this approach so that people would not have to carry an extra device around, which could be easily
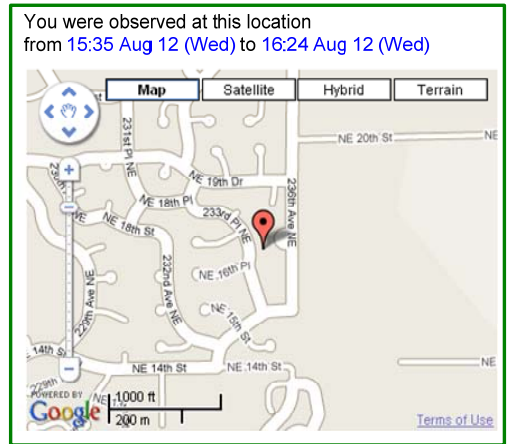
(a)

(b)

Figure 1: (a) Maps with timestamps reminds participants of the locations they visited. (b) Participants were asked to answer a set of questions for the places regarding to sharing location with people in four social groups (i.e. family member, close friend, acquaintance, and stranger).

forgotten at home or work. The location sensing application was previously developed by Benisch [13], and was run continuously in the background using both GPS and Wi-Fi positioning. The phone's geo-coordinates were recorded every 15 seconds if the embedded GPS unit was able to determine its position. Otherwise, the application recorded visible WiFi MAC addresses every 3 minutes instead. All these readings were stored in a file on the phone.

Each day, we reminded participants to upload their location trace to our server, via a connection to a personal computer. We used this approach since most of our participants did not have a data plan on their SIM cards. Afterward, participants were asked to log onto our web application and answer questions about the places they visited.

When participants uploaded their location file, our web application automatically translated the Wi-Fi AP addresses into geo-coordinates using Skyhook's API[8]. Our web app then parsed the traces and identified salient places, based on places participants stayed for more than 5 minutes. Our web app then displayed a map (see Figure 1a) showing visited places with corresponding timestamps, to remind participants of where they went. Participants also answered questions

about sharing location information with people in four different social network groups (i.e. family members, close friends, acquaintances, and strangers). We collected data about the first three of these groups in an entrance survey, and used names of people provided by participants.

For example, in Figure 1b, "Mary" is the name randomly drawn from this participant's family members. This participant was asked to imagine the scenario in which her family member "Mary" would like to know her location. The participant then responded to the following questions:

- How comfortable she (this participant) would be to let "Mary" know where she was at the specific moment.
- How familiar "Mary" was with the place.
- Terms or phrases she would like to use to refer to this location in the specific situation.

For strangers, participants did not see the question regarding the other party's familiarity with a certain place.

To provide more confidence that our results could generalize, we recruited participants from a university that has a presence in two different cities, one (city A) on the east coast of the U.S. and one (city B) on the west coast. We posted flyers around both campuses, and advertised on university mailing lists. We recruited twenty-six students (12 female) ranging in age from 20 to 44 years old (mean=25.6, SD=5.8). The students had a diverse range of majors, with 18 participants coming from city A and 8 from city B. Of the 26 participants, eight of them traveled outside the city they live in while the study took place.

Participants were compensated with a $30 USD online gift card. No real location sharing took place in our study.

4 CLEANING THE COLLECTED PLACE DATA

After collecting all the data from our participants, we cleaned up the data in three ways: filtering out irrelevant entries (less than 2% of total records), deriving extra attributes (see following section), and labeling each place name with category information (described shortly below).

4.1 Filtered Location Entries

We removed some entries due to positioning error (less than 0.5% of all the records, based on daily feedback from participants on their location trails). Other entries were removed due to unlikely scenarios, (less than 0.5% of all records, such as sharing location with a family member when they were both at home). Entries without meaningful place names were also filtered out (less than 1% of the records, including, for example, empty strings, random characters, "n/a", "nothing", etc.).

After removing these entries, we had 118444 location readings from 26 participants. We extracted 403 unique places visited and 1157 distinct descriptions for these 403 places. On average, each participant visited 15.5 distinct places over the two-week period (median: 14, SD= 5.17).

Attributes	Explanations
(lat, lon)	Geo-coordinates of the place
FromTime	P's arrival time to the place
ToTime	P's departure time from the place
Group	The social group of R (Family member, close friend, acquaintance, or stranger)
PhyDist	The physical distance between P and R, in a scale of 1 to 4 (1=same city, 2=same state diff cities, 3=same country diff states, and 4=diff countries).
CmftShare	How comfortable of P letting R know where he/she was at that moment, in a scale of 1 to 7 (1= not comfortable at all, 7= fully comfortable)
Familiarity	How familiar R with this place, in a scale of 1 to7 (1=don't know this place, 7=extremely familiar. P can input "not sure" if they don't know the answer)
PlaceName	The place name which P would like to use in the specific scenario.

Table 1: Directly captured attributes, where P stands for Participants and R stands for Recipient.

Attributes	Explanations
DistHome	Distance from this place to P's home
DistWork	Distance from this place to P's work place
Duration	The amount of time P spent at this place
Freq	Number of times P visited this place
UserCount	Number of participants who visited this place
Entropy	The diversity of users visiting a particular place.

Table 2: Derived attributes

4.2 Derived Attributes

All the directly recorded attributes are shown in Table 1. We also derived some additional attributes from this data (Table 2), including, for example, the duration of each stay based on the arrival and departure time, and the distance from the target place to the participant's home and work location. Furthermore, based on aggregate data, we calculated how frequently a participant visited each place, how many participants in total have visited a place, and the entropy of a place (based on Cranshaw et al. [17]).

Location entropy characterizes the diversity of users seen in a particular place. Entropy can be used as a proxy for estimating how public a location is. That is, public places (like universities and cafes) tend to have higher entropy, while private places (such as homes) tend to have lower entropy. More formally, for a place visited by a set of participants U_L, the entropy is defined as:

$$Entropy(L) := -\sum_{u \in U_L} p(u; L) \log p(u; L).$$

where *p(u; L)* is the number of times a particular participant visited place *L* over the total times the place was visited by all the participants. To make the entropy more representative, we calculated this value not only based on the location traces collected in our study, but also
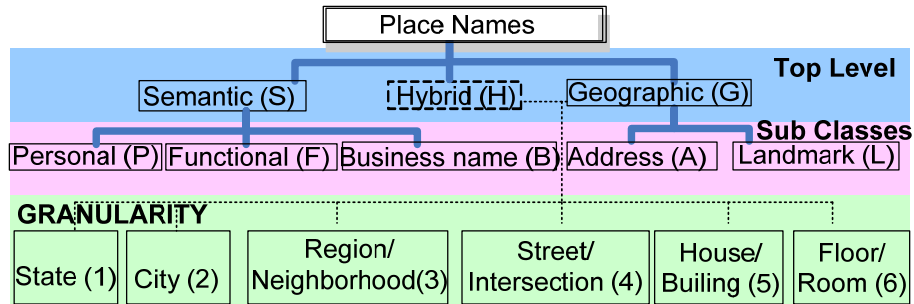
Figure 2: Place naming taxonomy. Semantic, geographic and hybrid naming are three top-level categories, and can be further sub-categorized into several classes.

combined with the location logs collected by Locaccino, a location sharing application [5, 17]. In total, over 2 million location logs were used in calculating entropy value in this work, describing the location traces of 493 users, each using Locaccino for a median of 38 days.

4.3   Place Naming Taxonomy

To understand people's place naming preferences better, we identified several patterns of how people name a place. Barkhuus et al.'s [11] proposed four types of location labels, namely geographic references, personal meaningful place, activity-related labels, and hybrid labels. We refined this classification by organizing these categories into a hierarchy with more fine-grained subcategories (Figure 2).

Based on the place names we collected, we saw that people used two major techniques for tailoring their location information. The first was to choose the perspective from which people address about the places, i.e. semantic, geographic, or hybrid. These perspectives are represented as *top-level categories* in Figure 2.

Semantic names can represent an official or informal name for a place, as well as its function. Examples include 'home,' 'coffee shop,' and 'Barnes & Noble.' Semantic names usually do not directly reveal the absolute position of a place, hence it might be difficult to pinpoint (or uniquely pinpoint) on a map without extra knowledge. *Geographic* names describe geographic locations, and include, for example, street addresses or nearby points of interest. Geographic names can usually be located at or near a specific point or area on a map. *Hybrid* names combine semantic and geographic information. Examples include 'Starbucks on Center Ave' and 'Barnes & Noble near Central Park.' Hybrid naming is often used to eliminate the ambiguity from using semantic information alone.

The top-level categories can be divided into 5 sub-classes: *Personal, Functional, Business name, Address, Landmark* (see Figure 2). The first three of these, personal, functional, and business name, are semantic names. *Personal* names refer to places that have highly personal meaning to individuals, such as 'home' and 'work'. *Functional* names reveal how a place is used and can imply what activities are carried out at those spots. Examples include 'restaurant,' 'gym,' and 'church.' *Business names* use the registered business name or trademark, such as 'Barnes &
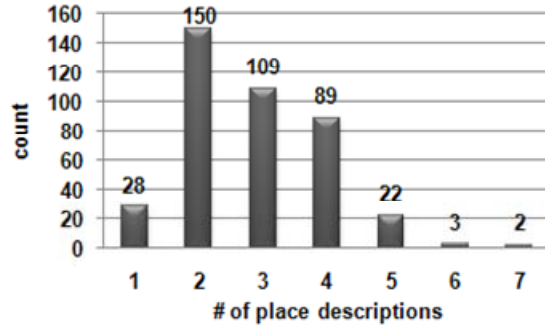
Figure 3: Histogram of number of place names for a given place. Among all 403 places in our study, 150 of them were associated with 2 different place names, 109 of them with 3 place names, and so on. The average number of place names associated with one place is around 2.8.

Noble' or 'Starbucks', to refer to the places. The latter two subclasses, address and landmark, are geographic names. *Address naming* uses the place's street address to describe the place. *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 second technique people used to tailor the location information was to tune the granularity of the disclosure, i.e. the precision of a disclosure. The precision can range from a large area to a specific spot. We identified a series of labels that corresponding to different granularities, which are shown in the bottom level of Figure 2. These granularities range from state level granularity to room level granularity. Here, granularity is only applicable to the place names that convey geographic information.

All the data collected from our user study were labeled according to the top level classes, sub-classes, and (where applicable) the granularity by two researchers. We computed Cohen's Kappa to cross-check inter-rater agreement of our labels. All three groups of labels had high agreement, i.e. $\kappa_{TOP} > 0.9$, $\kappa_{SUB} > 0.8$, $\kappa_{GRANULARITY} > 0.9$. The two researchers then discussed all the disagreed entries to come to a consensus on the final label.

5   DATA ANALYSIS

5.1   Observation 1 – Place Naming Diversity

As mentioned earlier, a single place can be associated with multiple names. This notion is supported by our data. On average, we saw *2.78* place names per physical location *(SD=0.89, Med=3, max=7, min=1)*. Figure 3 shows the distribution of the number of descriptions per place. About 39% of places had 2 names, 27% had 3 names, and 22% had 4. One participant even used 7 names to describe his work place to others, including 'office', 'at work', 'school', 'w building', 'x lab', 'y University', and 'z city'.
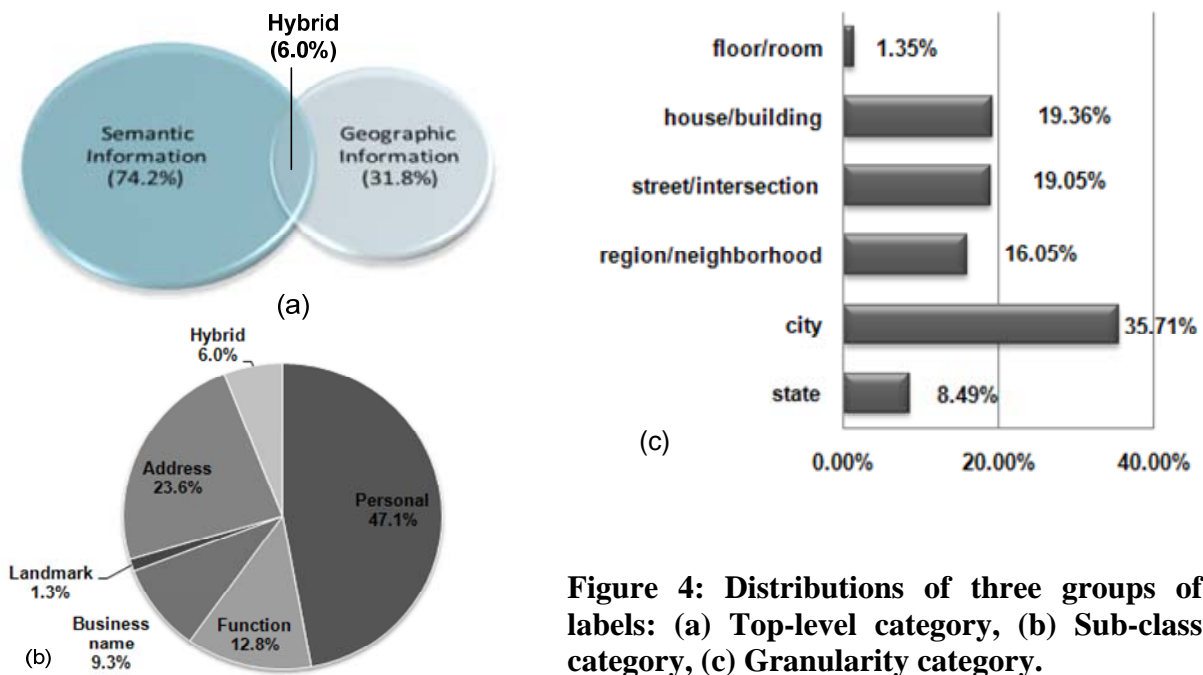
Figure 4: Distributions of three groups of labels: (a) Top-level category, (b) Sub-class category, (c) Granularity category.

Feedback from our exit survey suggested that using multiple place names was intentional and not due to inconsistency. People considered multiple factors when they decided what information they would like to disclose. As such, it was difficult for participants to find a single place name that was universally appropriate for all situations, and thus multiple place names were used.

5.2 Observation 2 – Information Blurring and Distilling

Consolvo et al. [15] claimed that participants did not intentionally blur their location for privacy reasons, i.e., disclosing something vague. However, our data suggests that blurring location is actually quite common and was used by people to modulate what information was disclosed, but in nuanced ways. We also observed our participants distill their location information into place names that emphasize the perspectives they want to share, such as inferring the functionalities of the places.

Figure 4 shows the distribution of each top and sub category. People used semantic information to describe their location most of time (i.e. 74.2%, Figure 4a). Geographic information is only used less than 1/3 of time. Among all the sub-classes, place names that describe personal places (e.g. "home", "friend's place", etc.) were used nearly half the time (see Figure 4b). We believe the wide use of semantic names is caused by the resultant force of both privacy and utility considerations. On one hand, semantic names might not be directly locatable, hence it gives people more confidence on their location privacy. On the other hand, semantic names distilled the underlying meaning of the target place which could significantly increase the utility of this piece of information.

In addition, among all the place names that contain geographic information, the histogram in Figure 4(c) illustrates the distribution of various granularities. Surprisingly, city level granularity appears most often. More than 79% of the time, these geographic names describe a vague region rather than a specific spot on a map. Therefore, by explicitly manipulating the granularity, people could blur their location to the degree they feel comfortable to share.

These observations suggest that when people have flexible ways to manipulate their location information, sharing their exact location directly is not preferred. For privacy considerations, instead of denying unwanted location requests, they tend to disclose something very vague to limit the amount of information shared. They also have tendency to distill useful information from their locations to make it easier for recipients to understand, hence the utility of the information could be guaranteed.

5.3   Observation 3 – Influential Factors

Researchers have already pointed out that people's privacy concerns and social relationships influence one's sharing behavior [15, 24, 25]. Our study confirms these findings and studies them in more depth. In addition, we also discuss two new attributes that haven't been examined previously: the recipient's familiarity with the place and place entropy.

Social Relationship: When we broke down these place naming methods by the recipients' social groups, we found that people used semantic naming more often when they had a close relationship with the recipient (see Figure 5a). To explain this phenomenon, we also plotted the distribution of place naming granularity in the same figure (right y-axis). When location was shared with more intimate social groups like family members or close friends, the portion of using geographic naming method was small (<15%) and the average granularity was finer (between street level and building level granularity). However, when the location information was shared with less intimate social groups, such as strangers, the usage of geographic naming was much higher but the average granularity drops dramatically (i.e. as coarse as city level granularity). This observation also confirmed people's location blurring intention get stronger when sharing with less intimate social group.

Comfort Level of Sharing: We also observed similar trend when we focus on people's comfort level of sharing. Figure 5b shows the distribution of the top level place name categories and granularities grouped by different comfort levels of sharing. In general, the usage of semantic place names goes up with the increase of people's comfort level of sharing their location. Furthermore, when people feel uncomfortable sharing their location (comfort level <3), they tend to use very coarse-grained geographic names (close to city level granularity in average.) When people feel extreme comfortable sharing their location (comfort level >5), although there only a small portion of time they use geographic place names, these place names reveal very specific position, hence are highly locatable. In Consolvo's work [15], the authors found that the usefulness of information was the primary reason for users to modulate the information to be disclosed based on participants' qualitative response in interviews. Our findings supplement their
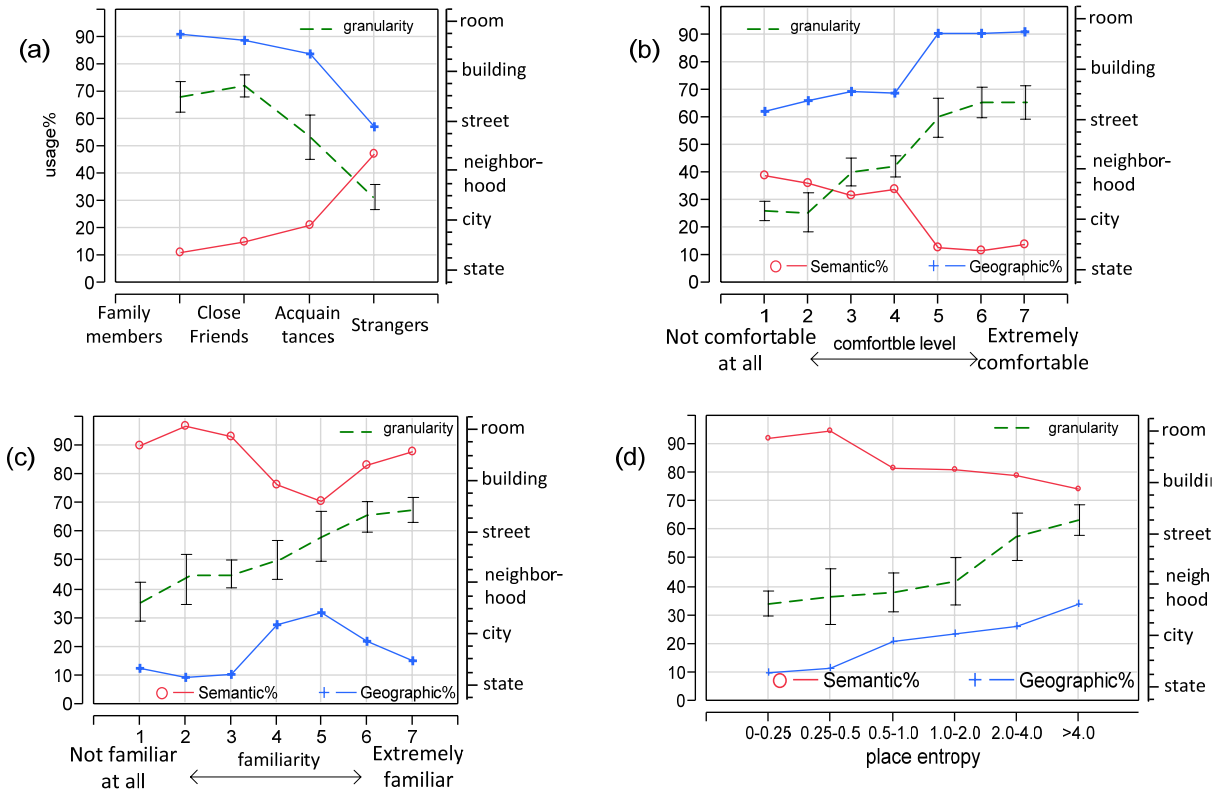
Figure 5: Important attributes that influence people's place naming methods (left y-axis) and place naming granularity (right y-axis), vertical bar indicated the 95% confidence intervals. (a) Sharing with different social groups; (b) Comfort level of sharing (c) Recipient's familiarity (d) Place entropy. The total percentage of semantic and geographic naming exceeds 100%, since some place names contains both of them (i.e. Hybrid).

observation with statistical data and suggest that people's level of comfort in sharing, a factor for location privacy, also plays an important role in determining what information to share, which in our case is what place naming method is used.

Recipient's Familiarity: With respect to familiarity, when people name a place, the literature suggests that people will consider how much knowledge they think the recipient has about that place, so as to provide more useful information[15]. Hence the recipient's familiarity with the place (in the sharer's mind) can influence the choice of place names. We grouped all the place names according to the familiarity rating, and measured the proportion of times semantic and geographic information were used (see Figure 5c). This plot suggests that the relationship between familiarity and the choice of place names is not linear. When the recipient is not familiar with the place (familiarity<=3), we saw that people tended to use semantic names, such as the function of the place. This finding makes sense since geographic information is not really meaningful to recipients unfamiliar with the area. For example, people shared names like

"grocery store" rather than provide the street addresses or neighborhood. When the recipient has some knowledge about the place, we observed an increase in sharing geographic information ($4 \leq$ familiarity ≤ 5). But when the familiarity gets higher (familiarity ≥ 5), the use of geographic names slightly drops.

On the other hand, if people do choose to name the place geographically, we observe the positive correlation between recipient's familiarity and the granularity of disclosure, i.e. people disclose more details of their position when the recipient is more familiar with this place, and vice-versa.

Place Entropy[1]: The other factor we examined is place entropy. A place with high entropy was visited by more users and is more likely to be a public place, and vice versa. For all the places our participants visited in city A, the average entropy value is 2.07 (SD=1.37, max=5.10, min=0.02847). We grouped place entropy into 6 intervals in base two log scale. Surprisingly, we observed a consistent positive correlation between the place entropy and the sharing of geographic information (see Figure 5d). Also, the granularity keeps on getting finer when the entropy increase. It suggests that people are willing to share more information about their absolute position when they are in public places. It could also indirectly suggest that people have less privacy concerns when they are in public.

All these observations illustrate the dynamics and the complexity of people's place naming preferences. They also give us important clues of how to model users' preferences.

6 DATA MODELING

In this section, we present the performance of our machine learning model. The characteristics of our data set motivated us to use the J48 decision trees in learning the place naming classifiers, which is able to capture the nonlinearity of the features, interaction between features, and can handle categorical and numeric attributes smoothly. Here, our goal was to see if we could predict the desired categories that people would use when naming in a given situation. As such, we do not solve the place naming problem entirely, but rather take a step towards doing so with this approach.

Given a participant *p*, learning from *p*'s own history could yield a very accurate model since people usually behave in routines. However, the concern here is overfitting. That is, we want to develop a generalizable model rather than one that is too specific to a given individual. Therefore, we separated the testing and training data so that no user appears in both sets at the same round. For each round, we randomly picked 5 participants (about 20%) for testing, and used the remaining data for training. We averaged the testing accuracies over the first 50 rounds (Table 3).

---

[1] In this sub-session, we only use data from city A to analyze the impact of place entropy. Since the data source ("Locaccino") for the entropy calculation doesn't have enough coverage in city B, hence the entropy values for places in city B are not as representative as the ones in city A.

	Avg Accuracy %	STDEV%
Top level category	85.5	3.14
Sub-Class	60.74	1.50
Granularity	71.25	3.44

Table 3: Average accuracy of predicting top level categories {semantic, geographic, hybrid}, Sub-class {personal, functional, business name, address, landmark}, and Granularity {state, city, region, street, building, room} labels.

The prediction of the top level class {semantic, geographic, hybrid} yielded an average accuracy of 85.5% (SD=0.03), the granularity prediction yield an average accuracy of 71.25% (SD=0.03), while the prediction accuracy of sub-class labels {*Personal, Functional, Business name, Address, Landmark*} is about 60.74%. When look at the prediction results in more detail, we found that many mispredictions (10.3% of testing set) happened between business names and functional place names. Many reasons could possibly contribute to these prediction errors, after going through these mispredicted instances, we found that given the same recipients and same locations, people were not always consistent in using functional naming and business names to describe these places. For example, people can use place names such as "Starbucks" (business name) or "coffee shop" (functional) interchangeably in real life, but a machine learning algorithm would take it as a mispredicted case. As such, we will need to take into account people's level of tolerance when we evaluate prediction results in future studies.

6.1 Effect of the Number of Days Included in Training Set

People might argue that two-week's data are not sufficient to build a prediction model. To validate our learning results, we analyzed the impact of the amount of data to the prediction accuracies. Here, we varied the amount of data included in the training set from 2 days to 14 days (the study lasted 2 weeks in total). Figure 6a shows how the average prediction accuracy changes with the amount of training data. We observe that the accuracy increased dramatically when the number of days gets larger at the beginning ($\leq$6 days). However, after one week ($\approx$8 days in the figure), the accuracies have a tendency to plateau.

This finding is explainable, since most people behave in routines. A week's duration that includes both weekdays and weekends could capture most of their routines. Hence, we see that the accuracies don't benefit a lot when more than 7 day's worth of data are used for training. In other words, at least a week's worth of history data is necessary for us to build an acceptable model, with additional data providing useful but smaller returns.

6.2 Effect of the User Profile

We are also interested to see whether we can boost the prediction results by carefully selecting the training set. The intuition is that people might have diverse preferences, such that for an individual participant *p*, a more accurate model could be built if we choose training sets that
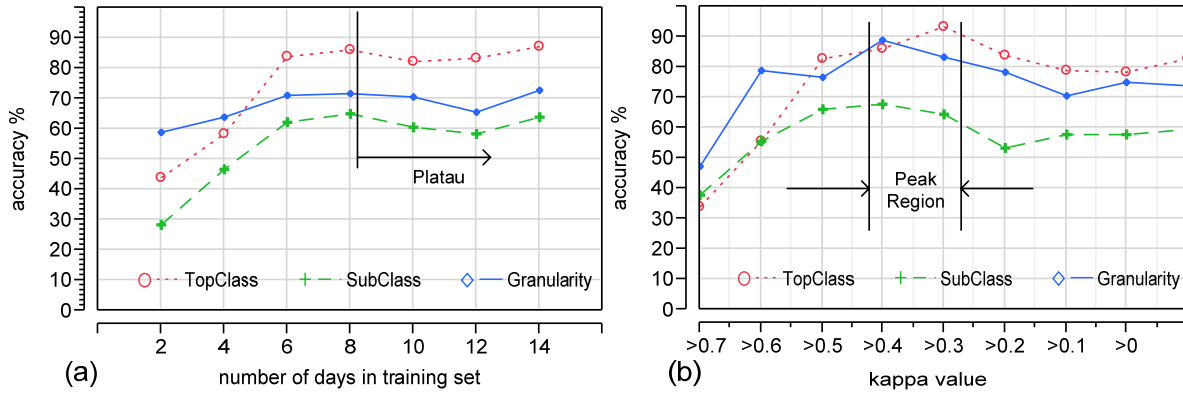
Figure 6: (a) Effect of the number of days included in training set: accuracies converge after one week. (b)Effect of grouping similar users: the highest accuracies appears when we group users with kappa larger than about 0.35

contain other people with similar preferences. In other words, rather than having a single general model for all people, we might have clusters of models.

In the entrance and exit surveys, we collected participants' demographic information and probed their sharing preferences by asking them to rate the level of comfort and usefulness in sharing place names to different social groups. We used these preferences and demographic data as user profiles, and estimated the similarity among all user profiles by computing pair-wise Fleiss' Kappa. For each participant p, this calculation lets us choose training data from other participants with similar profiles (i.e. who have Kappa value larger than k). We varied the value of k from 0 to 0.7 to see how it affected the prediction accuracy (see Figure 6b). The accuracies reach their peaks when the k values are close to 0.35. Thus, by grouping similar participants with k value around 0.35, we can achieve best performance in terms of the prediction accuracy when compared with the method of randomly separating training and testing data. With this approach, the accuracies of prediction for top-level class, sub-class and granularity labels are boosted to 93.2%, 67.8% and 88.7% respectively. We also observe that when the k is large (> 0.6) the accuracies are very low. Two reasons could attribute to this low accuracy: (1) not enough training data, since there are only a couple of participants that are highly similar to each other (kappa>0.7) in our dataset; (2) the similarity among user profiles could not fully capture the similarity of people's real place naming preferences. In other words, people with highly similar profiles might have different place naming preferences.

Although the user profile (demographic info and preference probing questions) we used to estimate users' similarity might not be optimal, it provides us insights that smartly choosing training set could potentially boost the performance of our models. Future work could also involve designing a set of profiling attributes that could better estimate similarity among users.

7 DISCUSSION

7.1 User Study Caveats

All the participants in our study were from a university community. We made our best effort in diversifying the sample pool by selecting people from different disciplines. Although we didn't observe a strong influence from attributes, like age, gender or status, follow-up user studies with more participants and greater diversity would provide more evidence that our results generalize. Moreover, participants' location information was not actually shared, but people might not behave in the same way when location is actually shared.

We also did not capture the purpose of sharing in our study, which could dramatically change place naming preferences for some cases. For example, if late for a meeting, a person might want to share very fine-grained location information. An actual deployment of a real location-sharing system that features place name presentation might confirm and improve our findings to some degree. However, had we actually shared people's location, this would have led to challenges in recruiting enough participants together with their friends, the bias of short and unvarying labels caused by typing on mobile devices, more time in building the experimental platform, and introducing more variables that would have made the data harder to analyze. This would have been foolhardy for a formative study. As such, we opted to do a "Lo-Fi prototype" to understand this space before actually building a system.

7.2 Automatically Generating Appropriate Place Names

Our long-term goal is to build a system that can automatically generate appropriate place names based on real-time context. The study and findings introduced in this report is a first step towards this goal, but there are still many challenges remaining.

First, while our work helps us understand what category people prefer when sharing place names, more work needs to be done to automatically associate tags (from grassroots efforts and existing databases) to the categories we proposed. Second, many resources (such as whitepage.com, yelp.com etc.) only record the center of businesses and POIs instead of the boundaries. In early prototypes, we found that simply using the nearest POIs leads to poor results, with many false positives regarding one's location. Here, there needs to be more work mapping one's current location to the actual point correctly. It is likely that one's personal location history can help in this effort. Third, existing positioning technologies have errors from several meters to tens of meters, making it hard to guarantee that the geo-coordinate input is accurate enough for generating appropriate place descriptions.

8 CONCLUSIONS

Most existing location sharing applications present users' location information on a map. However, sharing location in the form of appropriate place descriptions can provide more meanings and accommodate users' preferences better.

We studied the information people want to disclose in location sharing through a two-week-long study with 26 participants. We identified two general ways for manipulating the information

shared. We also proposed a hierarchy for how people name places. We examined the impact of different attributes on people's sharing preferences, and found that the recipient's familiarity with the place and the place's entropy can greatly influence how a place is referred to. By applying machine learning techniques, we were able to predict place naming categories with an average accuracy of higher than 85%.

Our findings suggest that it might be possible to develop more useful location sharing applications where appropriate place names are automatically (or semi-automatically) modulated. In future work, we plan to explore additional dimensions that might influence place naming, conduct larger scale studies with more diverse sets of participants. Future work will also look at the design, implementation, and evaluation of a location sharing system which presents dynamically generated place descriptions.

9 ACKNOWLEDGMENTS

10 REFERENCES

[1]   "foursquare," http://foursquare.com/.
[2]   "Google Latitude," http://www.google.com/latitude.
[3]   "Gowalla," http://gowalla.com/.
[4]   "Helio," http://www.helio.com.
[5]   "Locaccino: A User-Controllable Location-Sharing Tool," http://www.locaccino.org/.
[6]   "Loopt," http://loopt.com.
[7]   "PlaceLab," www.placelab.org.
[8]   "Skyhook API," http://www.skyhookwireless.com/.
[9]   D. Ashbrook, "Learning Significant Locations and Predicting User Movement with GPS," in Proceedings of the 6th IEEE International Symposium on Wearable Computers, 2002.
[10]   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.
[11]   L. Barkhuus, B. Brown, M. Bell *et al.*, "From Awareness to Repartee: Sharing Location within Social Groups," in CHI, 2008.
[12]   L. Barkhuus, and A. Dey, "Location-based Services for Mobile Telephony: a Study of Users' Privacy Concerns," in INTERACT, 2003.
[13]   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.
[14]   B. Brown, A. Taylor, S. Izadi *et al.*, "Locating Family Values: A Field Trial of the Whereabouts Clock," in UbiComp 2007: Ubiquitous Computing, 2007, pp. 354-371.
[15]   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, 2005.
[16]   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.
[17]   J. Cranshaw, E. Toch, J. Hong *et al.*, *Bridging the gap between physical locaation and online social networks,* CMU-ISR-10-107, 2010.
[18]   P. Fagerberg, F. Espinoza, and P. Persson, "What is a place?: allowing users to name and define places," in CHI 2003.
[19]   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.
[20]   J. Hightower, "From Position to Place," in The Workshop on Location-Aware Computing, 2003.
[21]   J. Hightower, S. Consolvo, A. Lamarca *et al.*, "Learning and Recognizing the Places We Go," *UbiComp,* 2005.
[22]   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.
[23]   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.
[24]   G. Iachello, I. Smith, S. Consolvo *et al.*, "Control, Deception, and Communication: Evaluating the Deployment of a Location-enhanced Messaging Service," in UbiComp, 2005.
[25]   G. Iachello, I. Smith, S. Consolvo *et al.*, "Developing privacy guidelines for social location disclosure applications and services," in Proceedings of the 2005 symposium on Usable privacy and security, Pittsburgh, Pennsylvania, 2005.
[26]   C. Jiang, and P. Steenkiste, "A Hybrid Location Model with a Computable Location Identifier for Ubiquitous Computing," in UbiComp, 2002.
[27]   D. H. Kim, J. Hightower, R. Govindan *et al.*, "Discovering semantically meaningful places from pervasive RF-beacons," in Proceedings of the 11th international conference on Ubiquitous computing, Orlando, Florida, USA, 2009.

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

[29]   J. Krumm, and E. Horvitz, "Predestination: Inferring Destinations from Partial Trajectories," in UbiComp, 2006.

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

[31]   S. Lederer, J. Mankoff, and A. K. Dey, "Who Want to Know What When? Privacy Preference Determinants in Ubiquitous Computing," in CHI, 2003.

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

[33]   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.

[34]   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.

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

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

[37]   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, 2007.

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

[39]   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.

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

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

[42]   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.

[43]   J. Y. Tsai, P. G. Kelly, L. F. Cranor *et al.*, "Location-Sharing Technologies: Privacy Risks and Controls," *TPRC*, 2009.

[44]   J. Wang, and J. Canny, "End-user place annotation on mobile devices: a comparative study," in CHI, 2006.

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

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

[47]   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, 2003.

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

[49]   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.

[50]   C. Zhou, P. Ludford, D. Frankowski *et al.*, "An experiment in discovering personally meaningful places from location data," in CHI, 2005.

[51]   C. Zhou, P. Ludford, D. Frankowski *et al.*, "How Do People's Concepts of Place Relate to Physical Locations?," in INTERACT, 2005.

[52]   C. Zhou, P. Ludford, D. Frankowski *et al.*, "Talking about Place: An Experiment in How People Describe Places," in Pervasive, 2005.