Supporting Location-Aware Services for Mobile Users with the Whereabouts Diary

Nicola Bicocchi, Gabriella Castelli, Marco Mamei, Alberto Rosi, Franco Zambonelli
Dipartimento di Scienze e Metodi dell'Ingegneria – Università di Modena e Reggio Emilia
Via Amendola 2 – 42100 Reggio Emilia – Italy
name.surname@unimore.it

ABSTRACT

Modern handheld devices provided with localization capabilities could be used to automatically create a diary of user's whereabouts, and use it as a complement of the user profile in many applications. In this paper we present the Whereabouts diary, an application/service to log the places visited by the user and to automatically label them with descriptive semantic information. In particular, Web-retrieved data and the temporal patterns with which different places are visited can be used to automatically define such meaningful semantic labels. We describe the general idea at the basis of the whereabouts diary and discuss our implementation and associated experimental results. In addition, we illustrate several applications that can fruitfully exploit the whereabouts diary as a supporting service, and discuss areas for future work.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Systems**]: Distributed Systems; I.2.3 **[Artificial Intelligence]** Probabilistic Reasoning.

General Terms

Algorithms, Location-based Services

Keywords

Location-aware services, Localization algorithms, Bayesian networks .

1.INTRODUCTION

The recent diffusion of handheld devices and smart phones equipped with localization capabilities is opening new scenarios in the development of context-aware services. In this paper we present the ideas and a first prototype implementation of the *whereabouts diary*: an application running on a GPS-equipped handheld device that records the list of relevant places visited by the user. The diary runs autonomously without requiring user's interactions and is able to classify *semantically* the places being visited in an unsupervised way. Relevant places can be extracted by considering clusters and dropouts in the GPS signal (typically indicating the user staying in a place or entering in a building). Semantic information can be added by exploiting the structure of people daily routine. For example, the place where the user usually spends night-time can be tagged semantically as "home", while the place where the user usually goes from 8am to 6pm can be tagged as "work". Specifically, we realized a set of Bayesian networks to diagnose the kind of place given the temporal pattern of user visits. Further information can be extracted by geocoding the place and mining the Web in search for relevant information.

The result is a diary describing the user daily life and that could be used either as a stand-alone application allowing a user to browse through his past locations, or most importantly as a supporting tool for other applications that could take advantage of such a profile information.

From the latter perspective the Whereabouts diary can be regarded as a sort of middleware-level service to support location-aware mobile services. While most of the proposed middleware infrastructures focus on "low-level" issues like communication and interaction among nodes, the Whereabouts diary focuses on high-level issues such as context-representation and its understanding (e.g., it encodes what places matter to the users, and what they mean to them). In our opinion, such kind of infrastructures will become more and more important in the future. As context information will become more detailed and relevant, there will be the need of managing them at the middleware level, providing application with a pre-digested and understandable view of such data [Cas07].

The remainder of this paper is organized as follow. Section 2 presents the main algorithms at the basis of the whereabouts diary and of its construction. Section 3 presents experiments and discusses performance of our implementation. Section 4 presents a tourist application taking advantage of the diary and also sketches some other applications that could be realized on top of the diary. Section 5 presents related work. Section 7 discusses areas for future work, and Section 8 concludes.

2.THE WHEREABOUTS DIARY

The construction of the whereabouts diary is an incremental process. In this section we describe *(i)* how to extract the places visited by the users in terms of their geographical locations, *(ii)* how to derive addresses and business' activities from such places, *(iii)* and finally how to add personal semantic labels (e.g., home, work, etc.) to the places.

2.1Diary based on Coordinates

At the most basic level, the diary is an application that continuously collects and stores the user location. In our implementation the diary collects a log of GPS reading sampled at 0.1 Hz. This creates a log like the one presented in Fig. 1.

Longitude	Latitude	Time

Mobilware 2008, February, 12-15, 2008, Innsbruck, Austria.

-73.974	40.763	July, 4, 2006, 4:35:00 pm
-73.973	40.766	July, 4, 2006, 4:35:10 pm
-73.974	40.765	July, 4, 2006, 4:35:20 pm
…	…	…

Fig. 1 Log of GPS coordinates.

Starting such a log of the GPS readings, it is possible to run segmentation and clustering algorithms to infer the places where the user spends most of his time [HigC05]. Following an approach similar to the one proposed in [MarS00, SchR06], we tagged as relevant those places for which either one of the following conditions apply:

1. the GPS signal is lost for at least T seconds and it is re-acquired later on at a distance of less than L meters from where it was lost. This reflects the situation in which a user enters a building and leaves it after some time. Some empirical evaluations let us to set T = 20 minutes, L = 20 meters. The constraint on time is important to wash out GPS signal glitches, the constraint on space is useful to avoid those situation in which the GPS has been shut down and the user moves away.
2. The GPS readings over a time window of W seconds are clustered within a radius of R meters from each other. This reflects the situation in which the user stays for a long time in a place like a park or a square. Some empirical evaluations let us to set W = 20 minutes, R = 100 meters.

The list of relevant places is built online and incrementally. When a set of coordinates meets one of the above criteria, the system looks in the list of the already discovered places for one closer than L = 10 meters to the coordinates. If such a place does not exist, a new place is created and the time of visit is recorded. If the place exists, the place coordinates are averaged with the new coordinates, and if enough time has passed since the previous visit (30 min), the time of the new visit is recorded.

The result of this operation is a list of places described in terms of longitude and latitude, and a list of time intervals associated to each of the coordinates indicating when the user has been there. This process creates a diary like the one presented in Fig. 2.

Longitude	Latitude	Time
-73.974	40.763	July, 4, 2006, 4:35pm-5:41pm
…	…	…

Fig. 2 Diary based on GPS coordinates.

2.2Diary based on Addresses and Businesses

A simple list of coordinates is only partially informative and the need of translating from positions to places (i.e., adding semantic meaningful tags to the discovered coordinates) has been widely recognized [Hig03]. A diary containing information like "*the user was at home*" rather than "*the user was at coordinates (10.873, 44.630)*" would be naturally much more informative and easy to use in context-aware applications.

A step in the process of adding semantic information would be to translate from coordinates to addresses. This can be done via standard tracking and geocoding services (as common GPS navigators do). However, because of errors in GPS localization and errors in the process of segmenting and clustering the GPS readings to identify relevant places, in most of the situations, it will not be possible to identify the unique address where the user is located, and only a partial estimate can be given. This second step converts the diary in the one depicted in Fig. 3.

In our implementation, the coordinates associated to places have been translated into addresses using a custom geocoding service. Most of the geocoding services available online (e.g., that provided by the Google Maps API) translate addresses into coordinates. Instead, the diary needs the reverse operation: from coordinates to addresses. We developed a "reverse" geocoding service for our region, on the basis of maps available from a commercial navigator software. To take into account GPS and geocoding inaccuracies, and the errors introduced by the place retrieving process, the diary application tries to reverse geocode all the addresses within a radius of 10m from the place being segmented. Thus, the diary actually creates a list of candidate addresses where the user has been.

Place	Time
123, 5th Ave, NY, USA	July, 4, 2006, 4:35pm-5:41pm
4,5,…,21, 26th St., NY, USA	July 6, 2006, 7:00am – 8:00am
…	…

Fig. 3 Diary based on addresses. Because of GPS errors multiple addresses can be associated to a single place.

A third step can try to mine the Web to identify what is in a particular address. The primary source of information in this context would come from yellow- and white-pages services. However, due to the aforementioned localization errors, this process will return in some of the cases a list of all the businesses performed in the geocoded addresses. Still, in some situations a single exact match could be retrieved like in the case of the user being in a big stadium or entering a big shopping mall. Even more semantic information could derive by searching relevant events that happened in that place at that time. For example, it could be possible to extract from the Web the fact that "the 4th of July parade" took place near the geocoded location at the same time the user was there. This process could create a diary like the one depicted in Fig. 4.

Place	Time
4th July Parade	July, 4, 2006, 4:35pm-5:41pm
126, 13th St., NY, USA	July 5, 2006, 11:00pm – 7:00am
Starbucks Coffee || Uno's Pizza	July 6, 2006, 7:00am – 8:00am
….	…

Fig. 4 Diary based on places. Because of errors in the previous phases, multiple businesses can be associated to a single place.

To perform this operation, in our implementation, we screen-scraped information coming from a widely used online white-pages service[1] in our region allowing to query for who is at a given address. This operation is trivially achieved using the tools provided by the htmlparser[2] software. In particular, each geocoded address belonging to a given place (as provided by the previous

1 www.paginebianche.it

2 htmlparser.sourceforge.net

step) is looked up in the white-pages and the corresponding business is retrieved. The result of this process is a set of entries labeled with the possible businesses found in that place. This translation process is not completely accurate, since several addresses are not listed in the white-pages (mainly due to privacy constraints). Still, the fact that most public businesses (like shops, etc.) are listed, while several private houses are not, allows to prune out a lot of unlikely addresses being discovered by the previous step. Private spaces like "home" – that are likely not to be listed in the white-pages – can be derived from the following analysis.

It is finally worth pointing out that the operations described in this subsection could be automated and improved given the availability of a complete spatial database like the one integrated into commercial navigator systems.

2.3Diary based on Personal Places

Given that the user activities are profiled in some way (e.g., the diary may know a priori that the user tends to stay at home at night), the diary application can give labels to places by looking at the temporal patterns in which places are visited. For example, the place most visited at night during weekdays can be meaningfully labeled as "Home". Such kind of analysis can be also used in combination with commonsense information [Liu04] to disambiguate between alternative retrieved places. For example, in the table in Fig. 4 the ambiguity among "Starbucks Coffee" and "Uno's Pizza" can be resolved (at least from a probabilistic point of view) in favor of the former, in consideration of the fact the place has been visited from 7:00 am to 8:00 am.

To implement this step, for each place being identified in the first phase, the diary creates a Bayesian network to analyze the temporal pattern in which the place has been visited by the user (see Fig. 5).The Bayesian network is composed of 4 nodes.

1. The *weekend* node represents a boolean variable used to represent whether a given observation takes place in the weekend or not. This node is always observed on the basis of the information stored in the GPS signal. This information represents the variability in people behavior between weekdays and weekends.
2. The *hour* node is a 24-values discrete node storing the time of day. This node is always observed on the basis of the information stored in the GPS signal.
3. The *kind of place* node is a discrete node modeling what a given place is. In our implementation, we try to classify among 5 different kind of places: home, work, restaurant (to indicate any kind of dining place), pub (to indicate any kind of evening entertainment), and disco (to indicate any kind of late-night entertainment). This classification is rather arbitrary, and each user of the diary should provide the kinds of place that best match his habits. This node is never observed, and is inferred by probability computations.
4. The *happens* node is a boolean variable expressing whether the user visits that place at that time. This node is always observed on the basis of the outcomes of the diary localization phase.

The role of the Bayesian network is to encode the routine of the user daily life. This is done by compiling the probability distribution associated to the fact that the user, in a given moment, is in a certain kind of place. For example, the probability of the user being at home during weekdays is depicted in the table in Fig. 6.

Similar tables can be created for other kind of places. In our current implementation, these tables are compiled by hand by the users that are asked to self-report the likelihood of being in a given kind of place at a given time. Such kind of data could be derived automatically also by a labeled trace of user's past whereabouts, using standard learning algorithms [PatL03]. Once the tables are filled in, basic inference operations in Bayesian networks will be used to derive the most likely kind of place given the visit pattern.

Specifically, when the diary previous phases identify that the user is visiting a place, the corresponding Bayesian network is retrieved, and the *weekend*, *hour*, *happens* nodes are set to their actual values (the *happens* node is trivially set to *true* to indicate that there is a visit). Then, the diary computes the probability distribution of the *kind of place* node. The newly computed distribution will be used as a prior for subsequent visits. This naturally allows evidences to add up, actually enabling the Bayesian network to classify the places on the basis of the visit temporal pattern.

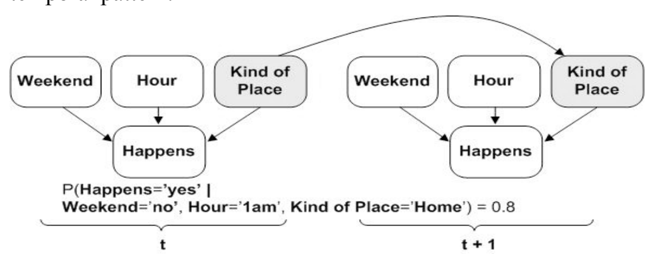

Fig. 5. Bayesian network to classify places. White nodes are those that will be provided as evidence.

Weekend = false, Kind of Place = home						
time	**11pm-6am**	**7am-8am**	**9am-1pm**	**2pm-5pm**	**6pm-7pm**	**8pm–10pm**
P(happens) = true	0.8	0.6	0.2	0.2	0.4	0.5

Fig. 6. Conditional probability table describing the probability of the user being at home during weekdays.

In its final form the diary represents a powerful source of context information allowing to extrapolate user's habits, preferences and routine behavior.

Of course, should other kind of sensing devices be available (e.g., RFID and NFC – Near Field Communication – readers), classification could use such information to better identify the places. For example, a powerful source of data could come from credit card transaction records that would identify not only in which shop the user has been, but also what he has bought (some recent proposals in the context of mobile wallet applications go in this direction [Fit04]). In the end, the combination of all the above steps leads to a diary close to the one in Fig. 7.

Place	Time
4th July Parade	July, 4, 2006, 4:35pm-5:41pm
Home	July 5, 2006, 11pm – 7am
Starbucks Coffee	July 5, 2006, 7am – 8am

….	…

Fig. 7 Diary based on personalized places.

3.EXPERIMENTS

To test the effectiveness of the whereabouts diary, we collected GPS traces for three weeks from three members of our research team (among the authors) as they went about their normal lives. Each member carried either an i-mate PDA 2K smart phone, or a HP IPAQ RX3700 pda, connected with a Bluetooth GPS reader. GPS signal has been acquired at 0.1Hz and processed on the fly by the handheld device. Overall, we acquired about 90000 GPS poses amounting at 360 MB of data. Overall, this resulted in 25 places being identified as relevant. During the data collection weeks, data collectors recorded ground-truth information about the places they have been. Such information has been collected with a simple notepad application running on the PDAs and allowing to write a textual description of where the user has been at a given time. In the following, we present some results obtained by comparing the whereabouts diary entries, after some weeks of usage, with recorded ground-truth information.

In a first set of experiments, we tried to verify the accuracy of the algorithm to identify relevant places on the basis of the GPS trace log. Following an approach similar to [HigC05], we classify the incorrect results into: *(i) wrong*: the user is in a place, but the diary reports he is in a different place, *(ii) false negative*: the user is in a place, but the diary reports he is moving, *(iii) false positive*: the user is moving, but the diary reports he is in a place. The results of this experiment are reported in Fig. 8, and they actually show the average of the results obtained by the data collectors. The results we obtained show that the algorithm is correct in 80.6% of the cases. This figure is coherent with the results presented in [HigC05] with regard to the A-S algorithm [AshS03] that is the one closer to our implementation. The high-percentage of false negatives (compared to the other cases) is mainly due to the fact sometimes the GPS takes a long time before acquiring the signal. Thus, it can happen that a user leaves a building, and the trace of the GPS is acquired only when he is already far away. In such a situation the place is not detected given the constraint on the maximum distance of spatial disconnection described in Sect. 2.1.

Correct	Incorrect		
	Wron g	False negative	False positive
80.6%	0%	17.3%	2.1%

Fig. 8. Errors in the algorithm to identify relevant places on the basis of the GPS trace log.

In a second set of experiments, we tried to verify the results of the (reverse) geocoding service. Basically, the idea is to verify the impact of localization errors in the process of geocoding. It is worth noticing that the maps we used to perform this operation record only the first and the last number of a street segment and span, uniformly, all the other numbers among the segment. This of course introduces further errors in that it does not take in to account the differences in the sizes of the buildings.

Since the place discovery algorithm clusters together points that are closer than 10 meters, we counted the number of addresses retrieved within a circle of 10 meters radius centered at the relevant place. The results of this operations are displayed in Fig. 9, and highlight two aspects. On the one hand, the address of almost half of the places can be retrieved uniquely (this is the case of large buildings – like the departments of our university). On the other hand, some places produce more than 10 associated addresses. This is the case of small buildings in the center of the city. It is fair to report that these distributions are rather preliminary since they are based on a dataset of only 25 places (those identified by the diaries of the 3 data collectors). We are currently conducting a more extensive data collection process that would allow us to identify more stable distributions.

In a third group of experiments, we tried to evaluate the performance of retrieving the businesses performed on a given place via the white-pages service. Comparing the results with the ground-truth annotations, we first tried to determine whether the correct place is retrieved. With disappointment, we verified that the actual place can be retrieved in only 40% of the cases. This is either due to localization or white-pages errors. Moreover, due to the multiplicity of addresses being discovered several businesses can be assigned to a given place. In Fig. 9, we report the distribution of the number of businesses found for a given place. It is easy to see that some addresses are not listed in the white-pages, since there are no places with more than 10 retrieved businesses. In addition, It is worth reporting that the number of businesses being retrieved is almost independent of whether the correct place has been found or not. Some places, in fact, return a long list of candidate entries not containing the correct one. The main source of errors of this phase is related to the white-pages interface and how it handles street numbers.

In our future work we will try to improve this result in many directions. First, we will try to integrate our software with commercial spatial databases (e.g., http://www.tomtom.com/pro). This would notably improve the coverage of the addresses and businesses being mapped. Second, we will try to implement more advanced localization algorithms in order to reduce the uncertainty about user location [Hig05]. Finally, we will try to embed commonsense reasoning (e.g., the user does not go to a disco in the morning) to cut off unlikely possibilities. The last resort would be to ask the user. In uncertain situation the diary could pop up and let the user solve possible ambiguities.

Finally, the last set of experiments verified the results of the Bayesian classification. Overall, our approach classifies the places correctly in 64% of the cases. In order to better analyze the results we tried to assess the confidence of the diary in its own classification – most probable estimate (MPE). To this end, we compute the information entropy of the resulting distributions. The lower the entropy, the more the system is confident about the MPE (i.e., the distribution peaks on the MPE value). More in detail, we separated the entropies related to the distributions that produced a correct MPE from those that produced a wrong MPE. For each of these two categories, we averaged together the entropies of the distributions producing the same MPE (see Fig. 10). Looking at the graph it is possible to notice that entropies related to wrong MPEs are higher than entropies related to correct MPEs. This is good and reflects the fact that the distributions associated to wrong estimates are less peaked, and thus the diary is less confident about its own classification. In fact, examining the wrong distributions in a lot of circumstances the distribution is bimodal: one peak is the correct one and another slightly more probable is incorrect. In such circumstances, it is likely that more observations on the temporal pattern of visits will correct the classification outcome.

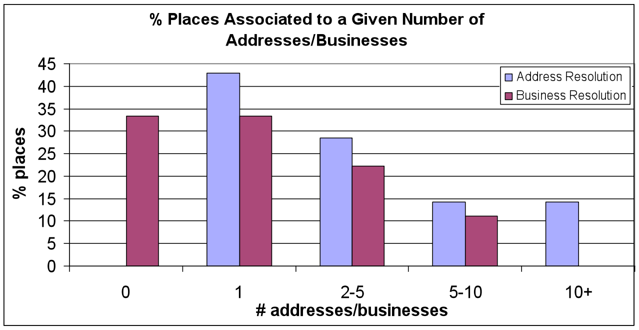

Fig. 9. Percent of relevant places corresponding to a given number of addresses/businesses.

Another interesting remark about the result in Fig. 10, relates to the average entropy associated to the different kind of places. Not surprisingly, *home* and *work* places have lower entropies since their associated temporal pattern of visits is defined more precisely. On the contrary, places like *pubs*, *restaurants* and *discos* have a more flexible pattern of visits and thus they are classified less precisely and tend to produce higher entropies.

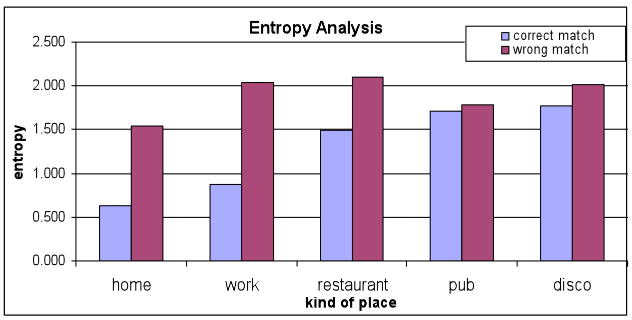

Fig. 10. Information entropy of the resulting distributions for the various kind of places. The information entropy for a 5-value discrete distribution ranges from 0 to 2.32 (flat distribution).

4.Applications

Once the diary is available, it can be used as a complement of the user profile in several application. In this section we present several application that could be realized or improved by connecting to the information extracted by the whereabouts diary.

4.1Tourist Diary

Tourist scenarios have always motivated ubiquitous computing applications. These kind of applications being deeply related to user profile and context-awareness can naturally take advantage of the whereabouts diary. An application could match the places visited by the user across a database of tourist attraction and suggest possible next visits. In addition, the exchange of diaries among tourists could allow to exchange experiences, hints and recommendations. For example, tourists traveling northward could fruitfully exchange diaries with backpackers traveling southward to obtain information about where they had just been [AxuV06]. Also in this case, the extraction of semantic information (e.g., retrieving shops and other businesses in a given area) could provide a notable added value to the information being communicated.

To test the whereabouts diary, we started implementing a guide application of our city offering advices and recommendation to tourist on the basis of their past travel pattern. For example, if the diary indicates that the user likes visiting museums, the guide suggests visiting other museums of a city.

The guide has been implemented as a Web application, based on Google Maps and accessible via PDA and smart-phones. The application woks on the basis of the following points:

- The list of all the attraction in the city is stored in a database also containing the typology of such attractions (e.g., name="Raffaello" type="cinema").
- The diary collects the list of places visited by the user and uploads it to the server.
- For each place identified by the diary, the list of all the tourist attractions within 80 meters is retrieved. Each attraction is associated with the probability that it was the one visited by the user (we simply calculate it with: 1 / number of retrieved attractions for that place).
- This information is stored in a database associating to each place all the retrieved attractions together with their type and with their probability (see Fig. 12).

Date	Duration	name	type	prob.
28-9-2007: 19:00	24000	Ariosto	cinema	0.33
28-9-2007: 19:00	24000	Il Pozzo	restaurant	0.33
28-9-2007: 19:00	24000	Raffaello	cinema	0.33

Fig. 12. database associating places with attractions together with their type and their probability.

Once a user requires a recommendation to the guide, the guide computes the most likely kind of attraction the tourist may be interested in. In particular, for each typology of attractions (e.g., cinema, museum, squares) the diary computes a score by adding all the time the user spent in a given kind of place multiplied by the associated probability:

$$score_{kind_i} = \sum_{places_{of_{kind_i}}} time_{of_{visit}} \cdot probability$$

The kind of attraction with the highest score wins and is proposed to the user. More specifically, the guide proposes the list of attractions of the winning kind ordered by the distance from the user and excluding those that the user has already visited (i.e., having a probability > 0.7 in the table in Fig. 12).

We are currently evaluating the performance of our guide application in terms of precision and recall. First results indicate that the whereabouts diary provides a rather effective support to the identification of places the user is interested in.

4.2Other Applications

Several other applications could benefit from the information contained in the Whereabouts diary. In the following of this section we present a list of such applications emphasizing how the diary could be useful.

Matchmaking. The automatic exchange of diaries between users can enforce powerful matchmaking capabilities. An application built on top of the whereabouts diary could try to match the places visited by the users to understand whether they are visiting similar places or not. The similarity between the places being visited can be a useful measure of the affinity between the two users. Similarly to what suggested in [KosN06], the diaries can be useful in the process of building common ground for face to face interactions. For example, a user could be notified that the person next to her has been to the same concert one week ago. With this regard, it is important to remark that the use of diaries enriched with semantic information is really useful in this context. For example, the application can automatically discriminate among colleagues (people spending time in the same work places), relatives (people spending time in the same home places) and friends (people spending time in the same pub and disco places). It would not be possible to conduct such kind of analysis on the basis of a purely geographic diary that only computes overlapping areas among people.

Pervasive Advertisement. Personalized advertisements offering users the most appropriate contents to suit their profile is a huge source of revenues for IT companies. The whereabouts diary could be a useful mechanism to apply personalized advertisement to ubiquitous computing scenarios. An application could show commercials to the user that are personalized on the basis of the diary. For example, if the user usually visits Au Bon Pain around noon, the application running on the user handheld device could present Au Bon Pain commercials at about 11am showing special offers, and indicating also those places that are in the neighborhood. In another setting, a wide screen display connected with a Bluetooth station could fetch the diaries of the people around and select the most appropriate commercials for the present population. In both these situations the whereabouts diary can provide useful information to compose the user profile. In addition, it is interesting to notice that the diary can provide useful feedback on whether the advertisement had been fruitful or not. For example, comparing the places visited by the user after seeing a commercial, it would be possible to infer whether the user followed the advertisement and actually went to Au Bon Pain.

Predict User Destination and Anomaly Detection. This application is based on the general assumption that people routinely performs repetitive actions (e.g., usually people go to work/school in the morning and come back in the evening, people have lunch around noon, etc.). On this basis, an application can learn from the whereabouts diary the user motion routine, and thus predict where the user will go next and eventually trigger alarms if anomalous deviations from the normal track are detected [PatL03, PatL04]. Such a service may be very useful for very young, elder or disabled people in order to support their independent living. For example, a child goes to school by himself; the application could monitor his movements alerting himself and his parents if notable deviations from the usual track are detected (the child may got lost or may have caught the wrong bus).

Personalized Navigation. MyRoute [PatC06] is an application to produce personalized navigation routes with the goal of reducing route complexity and cognitive burden. This is achieved by creating user specific routes on the basis of his previous knowledge about familiar routes and landmarks. In particular, MyRoute works by compressing directions coming from traditional navigation software into a single contextualized step (e.g., drive to work). The whereabouts diary could naturally support MyRoute by providing the information about users' relevant places.

Context-aware Instant Messaging. The messenger application Nomatic [PatX06] proposes to exploit places' semantic labels as meaningful status information in messenger applications. Such automatic labels could be used in combination of the standard ("on line", "not at PC", "busy" labels) to better express the actual status of the user. More in general, a wide range of ubiquitous computing applications can be enhanced by incorporating information about localization. For example, mobile phones can be automatically switched off inside a theatre, or could connect with the car speakers when the user is driving. Contact-management software could be programmed to automatically exchange business cards when the current place is recognized as work by both the users. Of course, the whereabouts diary could provide all these applications with suitable information.

Memory-aid. Memory-aid applications can take a notable advantage from the availability of the whereabouts diary. Specifically, notes, to-do-lists and reminders can be associated to specific places and be triggered by the user being in there. Moreover, the whereabouts diary can offer predictions on where the user is going next, so that reminders can be triggered also before the user reaches his actual destination. For example, if the diary predicts that the user will be close to the library the next day, the memory-aid application can remind the user to take the books on loan when exiting the house [AshS03].

Life blogging. Context-aware life blogging is like writing your personal diary in an automated way. The mobile applications Context Watcher [Koo06] and AniDiary [Cho07] run on smart phones, automatically connect to available sensors, log the information, and generate daily summaries about user's location, activities and moods. Also in this case, the whereabouts diary could provide high-level semantic location descriptions of the identified places. This of course could improve the life blog by making it more semantically expressive.

The broad range of applications where the diary could be useful supports our idea of having it as a general purpose tool offering important context information to third-party applications and services.

5.RELATED WORK

The recent availability of affordable localization mechanism and the recognition of location as a primary source of context information has stimulated a wealth of works addressing topics related to the whereabouts diary.

One of the earliest work trying to automatically compose a diary of users' whereabouts is the PEPYS application [New91]. This application uses IR badges and detectors to track user location in an indoor environment. On the basis of such an information, PEPYS compiles a diary of where the user has been and submits it to the user as a memory feedback. This kind of indoor localization systems, as well as its more modern incarnation (e.g., [Hah04]) could naturally complement the proposed GPS diary to deal with indoor settings.

The work described in [HigC05] compares three algorithms to cluster continuous GPS readings to find relevant places. Each of these algorithms could replace the clustering algorithm we described at the beginning of Section 2.1. However all these algorithms are only useful to spot relevant places and identify possible recurrent visit to the sample place. The problem of adding semantic information is completely neglected. The works in [AshS03, KanW04, MarS00, SchR06] presents similar kind of clustering algorithms.

The problem of adding semantic tags is posed, at least as an open problem, in [Hig03]. Other than clarifying the importance and the need for such a conversion from “positions” to “places”, the author illustrates two viable approaches to add semantics. The first approach is based on labeling places on the basis of the activities performed in there. The author proposes using RFID tags to infer users’ activities on the basis of the objects being touched (e.g., the user touches a fork and a knife, the system infers he is having dinner). Then, the system uses the activity (e.g., having dinner) to label the place (e.g., restaurant). The second approach involves humans assigning labels to places pro actively, and exchanging such labels among users. Neither of the two approaches has been actually realized, and they are mainly left as future work. In any case, once available, they could be well complement and integrate our proposal.

The works described in [PatL03, PatL04] adopt a Bayesian network to infer high-level user behaviors from low-level GPS readings. While their approach is similar to ours, their goal is different. While we try to classify the places, they try to classify user activities and eventually predict where the user will go next, on the basis of his past routes. It is worth to report that some user activities can directly identify the place in which they occur. For example, a sharp step in the user speed can reveal the user started/stopped driving the car. This automatically can be used to label that place as a parking place. Such kind of further information could improve the whereabouts diary.

6.OPEN ISSUES

There are several directions in which the presented work can be improved.

First, by exploiting data mining techniques and commonsense knowledge, one could think at more effectively discriminate among several candidate places [Liu04]. For example, if a person went to a restaurant at noon, it is very unlikely that he will go to another restaurant at 2p.m. As another example, a deeper knowledge of human activities could also suggest that a middle-aged woman is more likely to go to a supermarket after work instead of a young woman, the latter going to a pub with a higher probability. Due to these considerations, we envision an additional module able to learn habits and preferences of the users and automatically assign, in every situation, and with an increasing precision level, a probability value to all the candidate places. As the time comes forward, every user will tune his own model with observations coming from the real life. By this way, after a training phase, the system will gain accuracy and will fit itself over the peculiarities of every user. [Pat03]

A well-known and annoying drawback of this kind of systems is the training phase. To reduce the effort required by users to tune the system over their habits we propose an additional distributed layer, which can be considered a sort of social knowledge repository. We envision, in fact, to collect and aggregate thousands of users profiles over a central location. Indeed, using data mining techniques, a synthetic profile of each kind of user could be generated. This kind of "socially derived profile" can be used as a starting point for each user of our service. For example a middle-aged housewife will not start with an empty profile, but will be provided with a profile already conscious about common patterns of her social group. An interesting aspect of this approach is its own intrinsic incremental mode of operation. In fact, in every moment, every user can upload his own updated profile to the central location. By this way each user experience will contribute to increase the accuracy of the "socially derived profile" of his own group. Moreover, this kind of approach is prone to privacy problems, because a user is not called to exchange with the others the full log of its diary, but only a small profile describing its general behavior. This is especially important in consideration of the fact that several applications described in the previous section involve the exchange of diaries among users. With this regard, it would be interesting to investigate the possibility of exchanging hash-based signatures of the diary entries, instead of the actual values to preserve privacy [KosN06].

The system described in this paper is mainly based on the verified assumption that the most of the users usually spend the most of time in a small number of key places (home, work, preferred pub) [Pat03]. Due to this, a future improvement of our life diary is related to the ability of inferring, with a good level of accuracy, not only well known places but also casual and less predictable locations. Some considerations have to be done at this point. The first one is related to the real usefulness of this approach. We are in fact investigating how the overall accuracy of our system will increase by recognizing places in which users usually spend a little amount of their time. Is the overall complexity increase justified by a higher accuracy degree in recognizing relatively rare events? It mostly depends on the metric that we decide to use. Probably, to properly evaluate this kind of systems, a cost-based metric should be used. We think in fact that every activity, and then every location, of each user, has different relevance. Due to this, the overall quality of the places recognition engine is not only related with the ratio between good and wrong decisions but also with the cost that a wrong decision could have in different situations. For example if a predictive system is built upon our diary log, a slight variation in the places visited by the user in the afternoon can cause very relevant errors, and then costs, in the prediction for the evening. Although these considerations, which are driving us in investigating more sophisticated algorithms, we are also concerned about the feasibility of this task. In other terms: are little fluctuations manageable? With which kinds of mechanisms? Can the aforementioned "social" approach be useful in this task? Answering these questions is worth of further researches.

Another important aspect to be investigated is the possibility of exploiting other kinds of sensing devices and algorithms to extract more information about the place. For example, Wi-Fi localization could naturally complement GPS localization and improve its accuracy. Another powerful source of positional data is GSM cell ID. It’s very diffused, it doesn't require any particular hardware, it is available also indoor but, unfortunately, it is less accurate than GPS. Although this not trivial limitation, we think that the most of the applications described in this paper can be developed with a little additional burden even using GSM as main data source. We aim to identify these applications and verify benefits and limitations coming from the GSM system.

Although Wi-Fi and GSM can be considered alternatives to the GPS, we would like to make our system be able to use also not only strictly location oriented services, but also other sensing devices belonging to different domains. For example, a powerful source of data could come from credit card transaction records that would identify not only in which shop the user has been, but also what he has bought. Moreover we can consider other common sensing facilities of modern mobiles. For example, using Bluetooth meetings we can detect the presence of friends. And, in general, is more probable that three friends together are sitting in a pub instead of inside of a bank. Other good information sources can be the phone working mode (silent, normal) and the ratio of missed calls. For example we can infer that the user is located in a noisy environment by the fact that although his phone is in normal mode he has missed several calls. Our goal is to increase the accuracy of our system using these additional data sources and develop new algorithms to manage them.

The last, but maybe more complex and fascinating, perspective of our system is to develop an additional module able, not only to detect the current location of a user, but also to predict future positions. We envision a system in which time is not strictly referred to the present, but can be chosen arbitrarily. By this way, a service asking for the location of the user in some time in the future will receive a prediction that will increase its accuracy as the time comes forward in a seamless way. To reach this goal we need a model in which common activity patterns coming from all the sensors highlighted before are joined with user location. By this way we aim to build a system able to mine common behaviors but also, using some sort of case-based reasoning, strange or rare patterns. This work, beside its scientific value, can lead to several relevant applications. For example, the model of each user, including his favorites places and habits, could be used as a portion of user's profile in every application that needs it. Pervasive Advertisement can be a good example. Another interesting application can be some sort of personal agenda which is able to inform the user that he is risking to fail some planned activities due to his current, and then predicted, behavior. Moreover we envision to build a system that, obviously considering security problems, is able to publish the predicted user position autonomously. Using such a systems, friends and family, can for example be informed about the user behavior in a fully automatic way.

7.CONCLUSIONS

The Whereabouts diary can represent a very useful middleware service to support location-aware activities for mobile users. While most of the proposed middleware infrastructures focus on "low-level" issues like communication and interaction among nodes, the Whereabouts diary makes available high-level knowledge about the context and the user activities, and could be fruitfully exploited in a variety of application scenarios.

Although the presented results are very promising, and the current implemented version of the Whereabouts Diary is already usable, we have also identified several interesting directions along which to improve it and enrich it with further features, as areas for future research work.

8.ACKNOLEDGEMENTS

Work supported by the project CASCADAS (IST-027807) funded by the FET Program of the European Commission".

9.REFERENCES

[Ash03] D. Ashbrook, T. Starner, "Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users", Personal and Ubiquitous Computing, 7(1):275-286, 2003.

[Axu06] J. Axup, S. Viller, I. MacColl, R. Cooper, "Lo-Fi Matchmaking: A Study of Social Pairing for Backpackers", International Conference on Ubiquitous Computing, Orange County (CA), USA, 2006.

[Ben03] B. Benatallah, Z. Maamar, "Special issue on M-services", IEEE Transactions on Systems, Man, and Cybernetics, Part A, 33(6):665-776, 2003.

[Cas07] G. Castelli, A. Rosi, M. Mamei, F. Zambonelli, A Simple Model and Infrastructure for Context-aware Browsing of the World, International Conference on Pervasive Computing and Communication, White Plains (NY, USA), 2007.

[Cho07] S. Cho, K. Kim, K. Hwang, I. Song, AniDiary: Daily Cartoon-Style Diary Exploits Bayesian Networks, IEEE Pervasive Computing, 6(3):66-75, 2007

[Fit04] M. Fitzgerald, "Your Digital Wallet", MIT Technology Review, 24 August 2004.

[Fla06] J. Flanagan, "An Unsupervised Learning Paradigm for Peer-to-Peer Labeling and Naming of Locations and Contexts", International Workshop on Location and Context-Awareness, Dublin, Ireland, 2006.

[Hah04] D. Hahnel, W. Burgard, D. Fox, K. Fishkin, M. Philipose, "Mapping and Localization with RFID Technology", IEEE International Conference on Robotics and Automation, New Orleans (LA), USA, 2004.

[Hig03] J. Hightower, "From Position to Place", Workshop on Location-Aware Computing, Seattle (WA), USA, 2003.

[Hig05] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes, "Learning and Recognizing the Places We Go", International Conference on Ubiquitous Computing, Tokyo, Japan, 2005.

[Kan04] J. Kang, W. Welbourne, B. Stewart, G. Borriello, "Extracting places from traces of locations", International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots, Philadelphia (PA), USA, 2004.

[Koo] J. Koolwaaij, "Context Watcher", http://portals.telin.nl/contextwatcher/index.aspx

[Kos06] V. Kostakos, E. O'Neill, A. Shahi, "Building Common Ground for Face to Face Interactions by Sharing Mobile Device Context", International Workshop on Location and Context-Awareness, Dublin, Ireland, 2006.

[Liu04] H. Liu, P. Singh, "ConceptNet: a practical commonsense reasoning toolkit", BT Technology Journal 22(4):211-226, Kluwer Academic Publishers, 2004.

[Mam07] M. Mamei, R. Nagpal, Macro Programming through Bayesian Networks: Distributed Inference and Anomaly Detection, International Conference on Pervasive Computing and Communication, White Plains (NY, USA), 2007.

[Mar00] N. Marmasse, C. Schmandt, "Location-aware information delivery with commotion", International

Symposium on Handheld and Ubiquitous Computing, Bristol, UK, 2000.

[New91] W. Newman, M. Eldridge, M. Lamming, "PEPYS: Generating Autobiographies by Automatic Tracking", European Conference on Computer Supported Cooperative Work, Amsterdam, Netherlands, 1991.

[Pat06] K. Patel, M. Chen, I. Smith, J. Landay, "Personalizing Routes", Symposium on User Interface Software and Technology, Montreux, Switzerland, 2006.

[Pat03] D. Patterson, L. Liao, D. Fox, and H. Kautz. "Inferring high-level behavior from low-level sensors". International Conference on Ubiquitous Computing, Seattle (WA), USA, 2003.

[Pat04] D. Patterson, L. Liao, K. Gajos, M. Collier, N. Livic, K. Olson, S. Wang, D. Fox, H. Kautz, "Opportunity Knocks: a System to Provide Cognitive Assistance with Transportation Services". International Conference on Ubiquitous Computing, Nottingham, UK, 2004

[PatD06] D. Patterson, X. Ding, N. Noack, Nomatic: "Location By, For, and Of Crowds", International Workshop on Location and Context-Awareness, Dublin, Ireland, 2006.

[Rom06] G.-C. Roman, R. Handorean, R. Sen, "Tuple Space Coordination Across Space and Time", International Conference on Coordination Models and Languages, Bologna, Italy, 2006.

[Sch06] F. Schmid, K. Richter, Extracting Places from Location Data Streams, International Workshop on Ubiquitous Geographical Information Services, Munster, Germany, 2006