MENDEL
Soft Computing

PROVIDING CUSTOMISED INFORMATION PANEL CONTENT BASED ON USER BEHAVIORAL PATTERNS

Ivo Pisarovic[a], David Prochazka[a], Dan Vybiral[a], Jana Prochazkova[b]

[a]Mendel University in Brno, Faculty of Business and Economics
Zemedelska 1, Brno
Czech Republic
xpisarov@mendelu.cz

[b]Brno University of Technology, Faculty of Mechanical Engineering
Technicka 2, Brno
Czech Republic
prochazkova.j@fme.vutbr.cz

Abstract: *Although mobile applications are commonly using user location and behavior to provide relevant content, public information panels usually lack the ability to adjust the content for a particular user or a group of users. Therefore, we focused on the development of information panels that are able, in combination with a mobile application, to collect anonymous location data about the users, identify key behavioral patterns and provide content that is relevant for the users in the panel vicinity. The key property of our solution is the anonymity of the collected information and privacy in general. The proposed algorithm consists of the data clustering and subsequent analysis. Described solution can be used in any public building or campus that the users visit regularly.*

Keywords: *pattern recognition, clustering, location based service, user experience, user interface.*

1 Introduction

The vast majority of popular mobile applications is currently based on personalized content. The content of different social networks, weather forecasts, map searches and even news are based on user preferences and user behavior. The content personalization can be based on explicit user action (e.g. selection of topic or following a hashtag) or on analysis of user's location, previous actions [1, 2]. Especially this second approach is crucial for most services. There are at least two obvious reasons.

Firstly, the user does not want to be bothered by unnecessary actions. Each user starts an application with a clear intent (to get some information, to entertain himself/herself etc.). Any action that is not directly related to this intent is taken as an obstacle [3]. Secondly, especially in last years, the users are very cautious about providing personal information [4]. The only information that is commonly provided by the users is their actual location. E.g. more than 90% of United States smartphone users provide location to obtain location-based information [5].

Therefore, we focused on the design of an approach that is able to analyze the user behavior purely on the basis of the user location timestamps and use it for providing user-based content on public information displays. The first part of the article describes our algorithm that is able to process user's locations and distinguish different behavioral patterns. The second part of the article outlines its application on smart public information displays. These displays provide information that is related to the users in their vicinity.

2 Analysing User Behavior Patterns

The amount of user-generated content on the Internet has been steadily rising in the recent years [6, 7]. A large part of the generated information comes from the users of devices such as smart-phones, tablets and recently also smart-watches and other wearables. These devices can collect various kinds of data, such as location, data about movement or physiologic data. Applications running on these devices can use this data to infer various information. This information can then be used for further analysis, which can be later used to provide better content for the user. Depending on the use-case, the analysis can be done locally (on the device itself), or sent to a remote service to more computationally demanding analysis. The methods of analysis vary depending on the collected data and also on the expected result – the extracted information. Usually, the collected data come in the form of time series. Depending on the form, they can be used as a basis for a classification problem, a pattern extraction problem, or a prediction problem, among others [8, 9].

Currently, machine learning algorithms are often used as a solution for classification problems. Many recent papers focused on using machine learning algorithms for classifying. Nonetheless, these algorithms are more suitable for classification of data such as images, audio files or large data sets of numeric values. These datasets can be easily sampled with already existing methods and can be used as an input for classification [10]. However, location data sets, organized into a time series, are not very suitable for a classification problem and therefore other approaches should be used [11].

Using various models for pattern recognition is more fitting, as the model can be adjusted and modified, depending on the available data and also designed according to the intended output. In recent research, various proprietary models are designed, based on their input data. For example, a *MAST*-based models are often proposed. These models are focused on classifying various values on a single data entry. For example, these values can be a movement, an action, a current situation (location) and time. By supplying or classifying these four properties, the user's behavior can be analyzed and patterns can be extracted, as this model accounts for transitions between single states [12]. Parts of the models can be omitted, or added, depending on the available data and requirements of the information extraction.

Another approach to behavioral analysis is to model the user's behavior by proposing a persona, or multiple personas. These personas are simple patterns, which can be expected for the users and are used for pattern detection in the collected data. Every proposed persona has multiple properties, which the collected data should contain and evince. The more the collected data matches the proposed persona, the more the user fits the persona. After the persona is modeled, a decision tree can be used [10] to evaluate if the supplied data represent the persona or not.

2.1 Data Clustering

The clustering is often used in many research areas [13, 14]. In our previous work [15], we used location data collected from a custom smart-phone application and also exported data already collected by the *Google Timeline* service for our analysis. The data contained location entries of the user's whereabouts during the day along with timestamps representing the time they were collected. The application and the *Timeline* service usually collects location entries based on the movement of the user. The more the user moves, the more data points are collected, therefore if the user stays at one spot for a longer period, only one entry is collected. Users usually have from 5 to 50 different location entries collected during their days.

A timeline was constructed for each analyzed user, representing his/her movement during the day. This was firstly done by simply ordering the collected data by their timestamps. However, to provide a better framework for the analysis, we combined simple location entries into so-called *places*, which the user visited. These *places* were created from multiple location entries with a close proximity (with some radius tolerance to compensate inaccuracies). The timestamp of the chronologically earliest location entry represents the approximate time and date when the user arrived in the *place* and equivalently the timestamp of the chronologically latest location entry represents the time and date when the user left the *place* they were visiting. Thanks to this, we acquired a list of *places* each user visited, along with information about the individual visits (namely arrival and departure times). Statistics about known *places*, containing information about the usual arrival and departure times of all users, could be then constructed. The algorithm of clustering consists of a *for* cycle with two iterators p_1, p_2 that go through the chronologically ordered sequence of the visits. The p_1 holds the active record that defines the beginning of a new cluster. The p_2 goes through the sublist of following records until the end of the cluster is found. Then, the p_1 is moved to the beginning of a new cluster (see Fig. 1).

We focused on analyzing the user's behavior based on their place of residence and their workplace. A persona representing the most typical user was proposed. This persona was modeled according to available collected and processed data. It was modeled according to users who during their usual day leave their home in the morning, spend a larger part of their day at their workplace and then leave for some entertainment in the late evening. Finally, they arrive at their home again, sometime in the evening or during the night. This persona was generalized according to the working day stereotype. Different personas can be modeled for different kinds of users and their expected behavior, depending on the particular application and available data.

The modeled persona was used to search for patterns in the acquired and processed data, now containing information about visited *places*, as described before. All *places* for each user were scored according to Sect. 2.2.

2.2 The Evaluation of Possible Home and Work Place

We got the list of visited places $P_k, k = 1, \ldots, n$ with timestamps in previous step. Let λ_k^H and λ_k^W are the real number descriptors that determine particular places P_k. The computation of these numbers deal with the visiting frequency and time duration of the given place. We set:
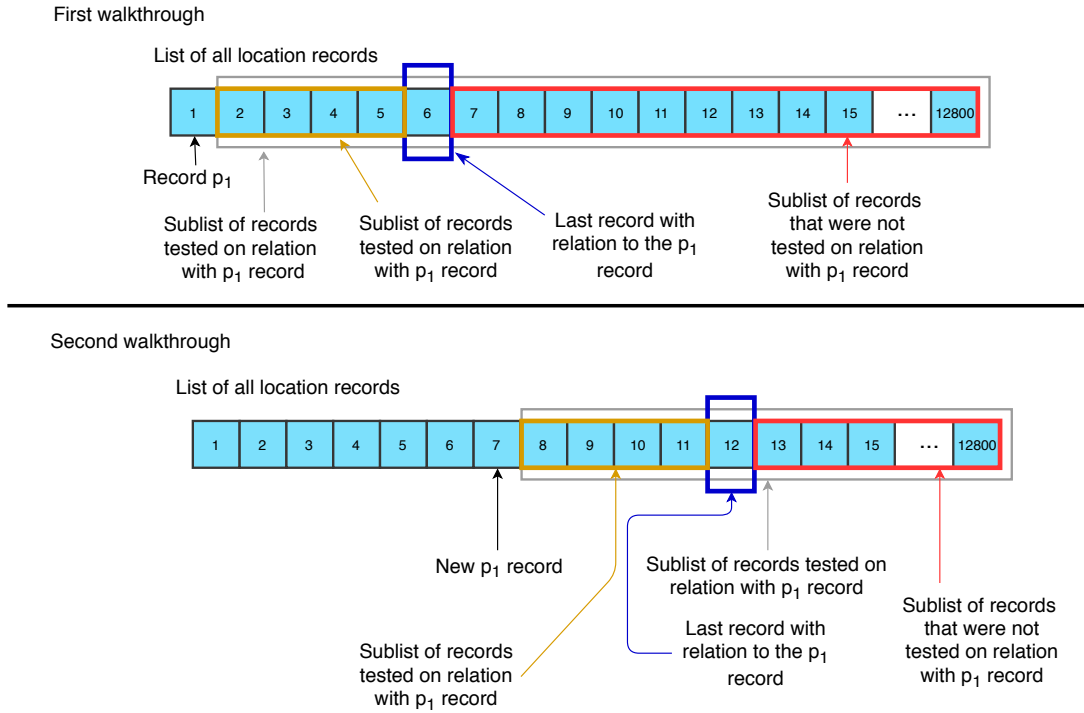
$$\lambda_k^H = 2h_k c_k p \tag{1}$$

Figure 1: Clustering of location records

$$\lambda_k^H = 2w_k c_k p \tag{2}$$

where the values $c_k, h_k, w_k, p \in \mathcal{R}$ are defined as follows. Value c evaluates the number of visits n_{P_k} in the place P_k in comparison to number of all visits in all places.

$$c_k = \begin{cases} 1 & n_{P_k} \leq \frac{1}{3} \sum_{k=1}^n n_{P_k} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The values h_k, w_k depend on the total number of arrivals and departures in day or night time of the given place P_k. Let $a_k^n, a_k^d, l_k^n, l_k^d$ be a values of the number of the arrivals in night, number of arrivals in day, number of departures in night, number of departures in day.

$$h_k = \begin{cases} 1 & \text{for} \quad a_k^n > a_k^d \quad \text{and} \quad l_k^d > l_k^n \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$w_k = \begin{cases} 1 & \text{for} \quad a_k^n < a_k^d \quad \text{and} \quad l_k^d < l_k^n \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The testing period was between two weeks and two years. Due to this differences, we set value $p = 1.5$ experimentally. This constant reduces the differences between time duration of data acquisition.

The maximal values of λ_k^H and λ_k^W detect corresponding possible residence place (home) and work place (P_{max}^H, P_{max}^W):

$$\lambda_{max}^H = \max_k \left\{ \lambda_k^H \right\} \tag{6}$$

$$\lambda_{max}^W = \max_k \left\{ \lambda_k^W \right\} \tag{7}$$

In the best case, one place of residence and one workplace would be detected, however, that is not always the case, as described further.

We tested this framework on data from 4 different users, who kindly supplied necessary data, or who approved and used our application for the collection of the data. Each user also had a slightly different lifestyle and day-to-day conduct.

The first user resides in the city of Brno and usually travels in the city's center. He/she usually arrives to work at some time in the morning and leaves irregularly during the afternoon. After work, he/she usually

visited different venues in the city, either for leisure or for other activities. He/she also regularly traveled to different Czech cities and vacationed in the various other countries. This user's data-set contains location entries spanning over 2 years. The correct workplace and place of residence were successfully extracted from his/her data-set, as can be seen in Fig. 2.
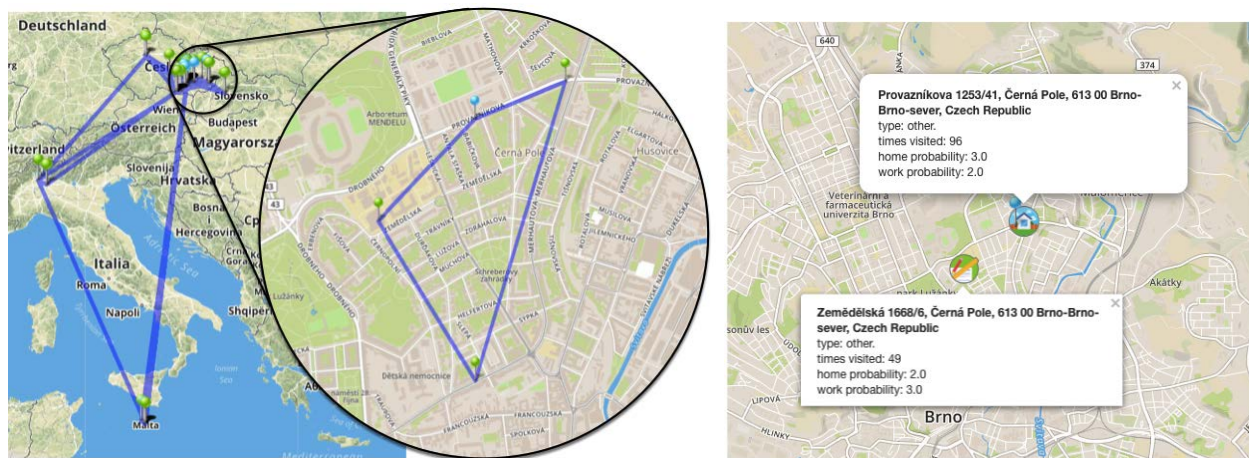

Figure 2: First user's movement data and extracted places

The second user behaved in a more consistent way, leaving home at similar times in the morning and also leaving the workplace at usual times. This user doesn't live in the city, but in a village nearby. Most of the visited places were only the user's home and workplace, with some different places visited across the country. The dataset contained location entries collected over two months time. In this case, both significant places were extracted successfully as well (see Fig. 3).

Figure 3: Second user's movement data and extracted places

The third user who supplied data behaved almost without any particular daily regime. The data was collected during one month. He/she traveled between his/her workplace and home highly irregularly. In most cases, he/she left home sometime in the late morning and left work at night, while in many cases arrived home in the morning hours. Also, this user visited a large number of different places with a smaller count of visits. He/she often traveled to more distant locations as well. The solution extracted the correct significant places in this case as well, despite the more chaotic data, as can be seen in Fig. 4.

The fourth and also the last data-set came from a user, who worked as a consultant and therefore did not have a regular workplace. He/she often visited various clients' offices, mainly in the city's center. This user also often traveled to their hometown, frequently not utilizing their home in the city. This data-set also contained entries collected only over two weeks. The number of visits of detected places was significantly lower than of the other users. One *place* was marked as both home and work location in this case, because other places had only a small number of visits (see Fig. 5).

3 Delivering Personalized Content on Public Displays

As mentioned, vast majority of personalized content is served via mobile devices. We want to enhance the user experience of public places by providing user-related information directly on information panels. Looking for

Figure 4: Third user's movement data and extracted places

Figure 5: Fourth user's movement data and extracted places

information on a mobile device can put a higher cognitive load on the user and it is generally a distraction. Hence, we use large display panels placed on walls at corridors as a complement to mobile devices. We needed to solve several problems related to displaying personalized contents on such devices:

- How to identify users in the display vicinity?

- How to resolve concurrency issues when different content should be displayed for different present users?

- Which content should be displayed on a mobile device and which one on a corridor display?

- How to exchange information between a mobile device and a display?

Firstly, we designed a method for identifying a user in the display vicinity (see Fig. 6). When a user enters the vicinity of a display, the user's mobile device receives a signal from a Bluetooth beacon placed near the display. The mobile device then knows which display is the closest to the user. Based on this information, the user's data and detected behavior patterns are sent to the server. Relevant contents are selected and displayed on the corridor display or on the user's mobile device.

Secondly, we need to solve the concurrency problem. For example, in the morning a lot of students are arriving to the university campus. Selecting the appropriate content can be tricky at such high frequency of moving users. We defined and implemented two basic rules:

- Contents on corridor display must be general enough. Personal content can cause data privacy issues. Moreover, content related to one person makes hard concurrency. The content should be general enough to meet requirements of more than one user.

- A priority is given to the contents. The priority can be furthermore raised based on the number of matching users.
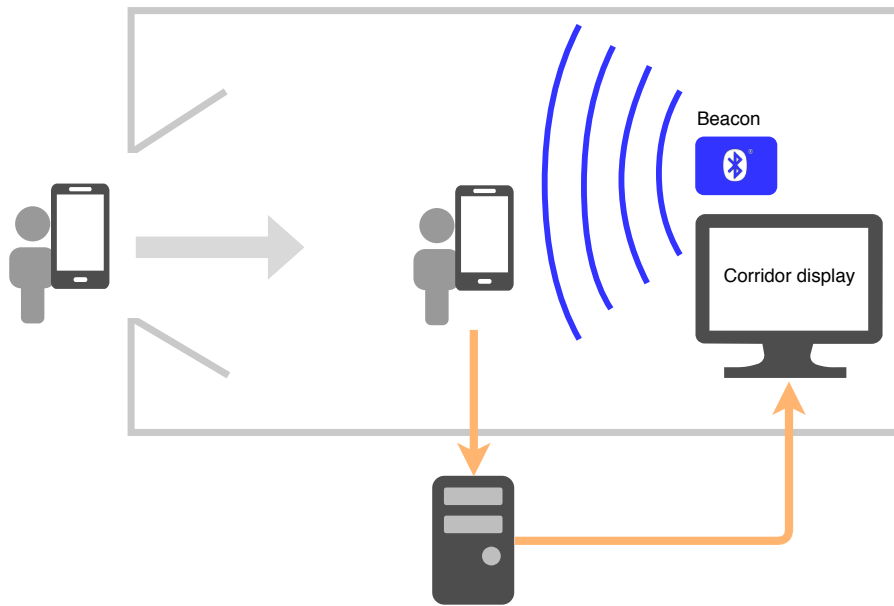
Figure 6: Identification of users in the display vicinity.

4 Personalized Information Use Cases

Based on the algorithm described in Sect. 2, we are able to distinguish places that are important for the particular user. The most important places are user's home and work locations. In addition, the same algorithm can be used at a more detailed level. When we enhance the localization precision by engaging beacons into the system, we will be able to monitor user's behavioral patterns even within a building. So, we are able to obtain a list of places that the user typically visits at certain times and days. We can even assume with high precision e.g. the courses user is visiting etc. The main benefits of this approach are:

- The approach is entirely automatic. The user does not need to manually fill the timetable, home address etc. That is very convenient for the user and enhances user experience. There is no initial configuration that can be discouraging.

- It is completely anonymous. No user login is required. Data are connected only with an anonymous user id (a number). A particular physical person cannot be simply identified.

- It can be used for predictions. We can predict users present at displays at certain times. Using this prediction, we can raise the priority of content that is relevant to these users. Content for typical users will be prioritized over content for current users. Therefore, the result will be a more stable concurrency resolving.

We applied our solution into the university environment. Our goal was to help the students as well as teachers to solve their common problems during their life at the university campus. You can see several examples in Fig. 7.

On the top left display in Fig. 7, you can see a table of forthcoming lessons, similar to the table of departures at airports. User's forthcoming lessons are automatically highlighted by the blue color at the display. It is worth to mention that a user does not have to set up the schedule because the user's lessons and classrooms are recognized automatically using our algorithm. However, even students or teachers without a cellphone can use this table. Their particular lesson is just not highlighted.

The top right display in Fig. 7 shows the nearest departures of public transport. Using the proposed algorithm we are able to automatically determine which direction the user is headed to (going home, to a gym etc.). Based on this, we can automatically highlight the best connection.

As mentioned, we are able to automatically obtain probable user's courses, the current year of study, the field of study, a favorite restaurant or favorite study rooms. Based on this additional study information, we can, for instance, display a notification for students of the last semester (see a diploma thesis deadline on the left bottom display in Fig. 7). Similarly, we can show typical questions and answers relevant to the needs of users in the display vicinity – e.g. if we detect that the majority of users in the display vicinity is in the first grade, we show them basic tutorials (like how to set up the Internet connection at the right bottom display in Fig. 7).
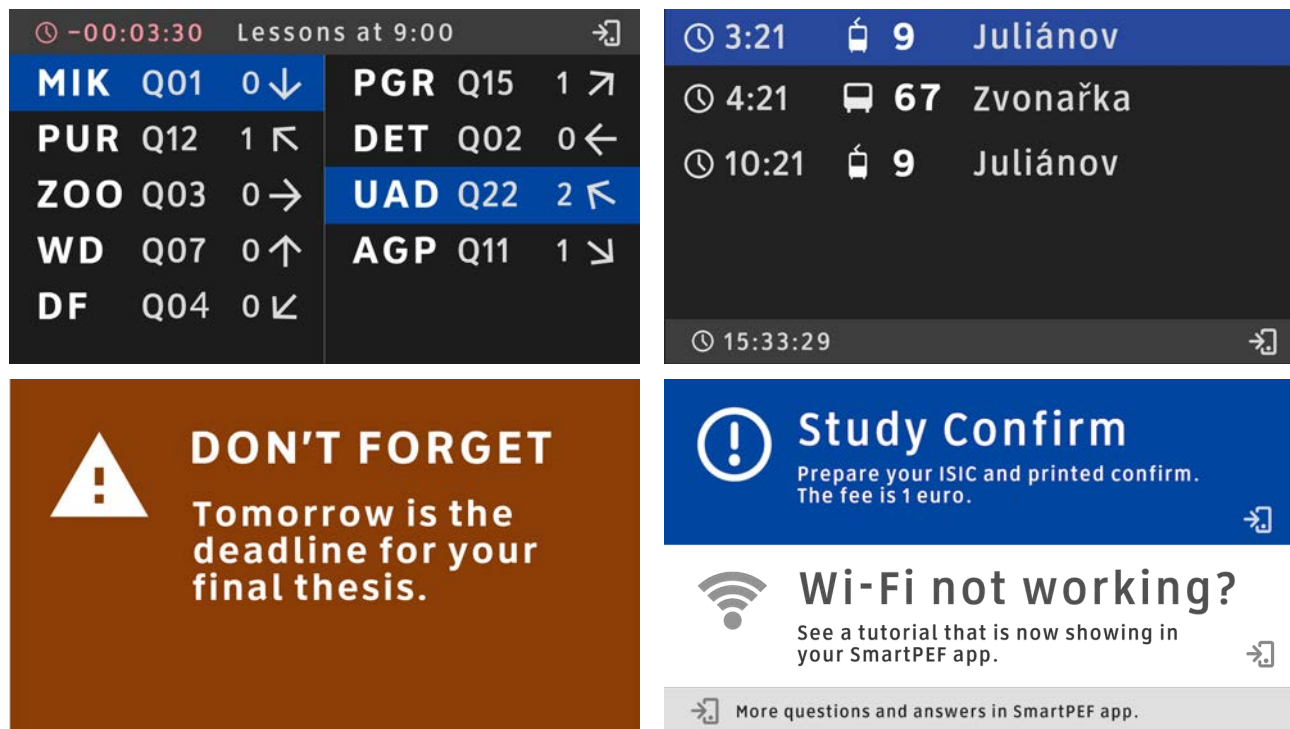
Figure 7: Examples of content displayed on corridor displays. The content is adapted to match needs of users present in the display's vicinity.

Such smart public displays radically change the way of providing personalized useful content to users. Showing information on a large corridor display is much more natural. It is a great benefit for the user. As a proof, try to compare looking for a bus connection on your mobile phone to simply raising your head and looking on a board with nearest departures. The second case is much more natural and comfortable. No explicit user action is necessary. It is completely automatic.

5 Discussion and Conclusion

The clustering algorithm described in Sect. 2.2 can be applied on datasets from the magnitude of buildings to the world-scale time series. The only difference is in the granularity of obtained information. It can be therefore used for a building as well as for city-wide solution. Interconnection with an external data source (e.g. classroom schedules) can be especially helpful. Such data sources can improve substantially our understanding of user behavioral patterns.

The whole system consists of several independent components. Thanks to this, the whole system is scalable and easily maintainable. For instance, simple web pages can be used for displaying the content on corridor displays. Hence, a common television or a monitor is sufficient. The presentation application can be shown in a built-in web browser on that television or using a minicomputer *Raspberry Pi* connected to a common computer display. As mentioned, a Bluetooth beacon is required to be placed at each display. For this purpose, battery beacons may be bought, e.g. *Estimote Proximity* beacons. Alternatively, you can cast the Bluetooth signal directly from *Raspberry Pi* minicomputer. The installation and configurations costs are therefore negligible.

The key part of the system is the server service for providing content. In our case, it is implemented using *Node.js* and in-memory document database *LokiJS*. It ensures low response times that are necessary for updating content on time. The used technologies are also good for prototyping such a new and experimental solution. However, it can be implemented using virtually any common server-side technology, however, the fast response is utmost important. It usually takes just a few seconds until the user leaves the display visibility area.

To verify the implemented solution, we have performed user experience testing. The test group was composed of university students not involved in the project. A set of simple tasks was given to the users (e.g. to find a particular lecture room). It was observed how they deal with the tasks. On the basis of the test results, several minor updates were implemented to the service user interface (positioning of the labels, changing time format etc.). However, no significant problems were detected. Despite the innovative and experimental nature of the system, users learned the basic principle of the system quickly without any help. They found the service very convenient.

References

[1] Steil, J., Muller, P.M., Sugano, Y., Bulling, A.: Forecasting User Attention During Everyday Mobile Interactions Using Device-Integrated and Wearable Sensors. CoRR, **1801.06011**. (2018) http://arxiv.org/abs/1801.06011

[2] Wang,Z., He,S.Y., Leung, Y.: Applying mobile phone data to travel behavior research: A literature review, Travel Behavior and Society. **11**, pp. 141–155. (2018) DOI 10.1016/j.tbs.2017.02.005

[3] Tidwell, J: Designing Interfaces, 2nd Edition, *Patterns for Effective Interaction Design.* O'Reilly Media (2010)

[4] Murphy, R.S.: Property Rights in Personal Information: An Economic Defense of Privacy, *Privacy*, vol. 2, second edn. Taylor-Fracis Group, London (2017)

[5] Location Based Services : Expected Trends and Technological Advancements In: *Geoawesomeness* http://geoawesomeness.com/expected-trends-technological-advancements-location-based-services/ (2017). [Online; accessed 25-May-2018]

[6] Lacy, S., Watson, B.R., Riffe, D. Issues and best practices in content analysis. Journalism and Mass Communication Quarterly **92**, pp. 791-811. (2015) DOI 10.1177/1077699015607338

[7] Naab, T.K., Sehl, A.: Studies of user-generated content:A systematic review. Journalism **18** (10), pp. 1256-1273 (2016) DOI 10.1177/1464884916673557

[8] Sjblom, M., Trhnen, M., Hamari, J.,Macey.J.: 2017. Content structure is king. Computers in Human Behavior **73** (C) 161–171. (2017) DOI 10.1016/j.chb.2017.03.036

[9] Smock,A.D., Ellison, N.B., Lampe, D., Wohn, D.Y.: 2011. Facebook as a toolkit: A uses and gratification approach to unbundling feature use. Computers in Human Behavior **27**(6), pp. 2322–2329.(2011) DOI 10.1016/j.chb.2011.07.011

[10] Geurts, P.:Pattern Extraction for Time Series Classification. In: Proceedings of European Conference on Principles of Data Mining and Knowledge Discovery. Lecture Notes in Computer Science. 2001, no. 2168 , pp. 115-127. Springer Freiburg, Germany (2001) DOI 10.1007/3-540-44794-6_10.

[11] Extracting Patterns using Neural Networks. In: *Stack Overflow* https://stackoverflow.com/questions/37639263/extracting-patterns-using-neural-networks (2016). [Online; accessed 25-May-2018]

[12] Song, J., Tang, E.Y., Liu, L.: User Behavior Pattern Analysis and Prediction Based on Mobile Phone Sensors. In: IFIP International Conference on Network and Parallel Computing. Lecture Notes in Computer Science. 2010, pp. 177-189. Berlin, Heidelberg: Springer Berlin Heidelberg, (2010) DOI 10.1007/978-3-642-15672-4_16.

[13] Jiang, S., Ferreira, J., Gonzalez,M.C.: Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. IEEE Transactions on Big Data **3** (2), pp. 208–219, (2017). DOI 10.1109/TBDATA.2016.2631141

[14] Kumar, K.V., Srinivasan, R., Singh, E.B.: A feature clustering approach for dimensionality reduction and classification. Advances in Intelligent Systems and Computing **378**, pp. 257–268 (2015). DOI 10.1007/978-3-319-19824-8_21

[15] Vybiral, D., Prochazka, D.: Identification of human behavior based on analysis of cellphone location data. In: Enterprise and Competitive Environment: Conference Proceedings. 2017, pp. 940-949. Brno, Mendel University in Brno (2017)