A General System for Self Collecting Individual Data - Application to Medical Data

Michel Schneider[1], Suan Tay[1], Chloé Gay[2], Marinette Bouet[1], Emmanuel Coudeyre[2]

michel.schneider@isima.fr, suan.tay@gmail.com, marinette.bouet@uca.fr, {cgay, ecoudeyre}@chu-clermontferrand.fr

[1]LIMOS, Université Clermont Auvergne, France
[2]Service MPR, CHU, Université Clermont Auvergne, France

***Abstract*-In this paper, we propose a system to allow the self-collection of individual data through digital questionnaires and sensors. The self-collecting person uses a tablet or a smart phone and wears a watch that contains the sensors. The main feature of this system is to memorize in a uniform way the answers to the questionnaires and the values of the sensed parameters in order to facilitate their joint analysis. The system has been developed and tested to monitor osteoarthritis patients. It represents an essential element of the control loop: elaboration of recommendations – monitoring the execution of these recommendations - evaluation and readjustment of these recommendations. This system has been implemented in a generic form and can be used to monitor any patient at home or on the move outside. We explain, also, the extensions we are currently making to obtain a general and flexible system.**

Keywords-self collection; questionnaire; sensor; mobile device.

I. INTRODUCTION

The initial motivation for this work resulted from the need expressed by the medical profession to have an easy-to-use system for the self-collection of health data. This involves checking that the patient is following the recommendations made, measuring the results of these recommendations, and then analyzing the results to readjust the recommendations. For example, in the case of the osteoarthritic patients we studied, it is a question of collecting the information relating to physical activity, difficulties encountered in performing certain movements, taking medicines, etc.. Some information can only be collected by questionnaire, but other pieces of information can be conveniently collected via sensors (e.g., the number of steps performed in a day). There is therefore a real interest in associating the self-collection of personal data by digital questionnaires and by sensors.

Self-collection by digital questionnaires has long been considered in all areas. Various systems have been suggested to create questionnaires and enable online responses. The best known are Google Forms and Lime Survey [1]. Questionnaire collection raises problems of relevance that are discussed in [2]|3][5]. In the medical field, questionnaires have been validated to collect various pieces of information about a state of health (see, for example, [4] for evaluating personality, [6] for evaluating level of anxiety), or a practice (see, for example, [7] for evaluating physical activity).

Numerous studies have focused on the collection of medical data by sensors. There is a wide variety of systems and sensors [8][9]. One of the challenges is to capture in a reliable and precise way physical activities [10][11] because they are an accompaniment to many therapies. Our goal was to develop a self-gathering system that combines the two modes of collection (questionnaires and sensors), and to store data collected in a uniform manner in a warehouse so that they can be manipulated jointly. It is thus possible to carry out analysis combining the two types of data in order to search for correlations or to compute indicators which, will serve to improve the recommendations for the patients. For example, for osteoarthritis patients, we can search correlations between the number of steps per day (data coming from sensors) and pain level or difficulty to make some movements (data coming from questionnaires).

Systems combining questionnaires and sensors have already been proposed, but for specific purposes [12][13][14]. The work of [12] uses questionnaires to collect the values of situational variables and wireless sensors to collect cardiac activity and physical activity. But, the two types of data are stored and processed separately. In [13], monitoring physical activity using wireless sensors is experienced and discussed. Drawbacks are highlighted and it is suggested to combine sensors and questionnaires. The work of [14] studies the effects of different treatments on the quality of life for adults with diabetes. This study is based on data coming from sensors and others coming from questionnaires. But, the two types of data are not integrated in a same system. To our knowledge, there is no proposition for an integrated system able to deal with the two types of data. The main advantages of our system are the following: uniform and integrated treatment of data coming from questionnaires and that coming from sensors, full mobility of the patient, management of the system by the medical staff itself without the intervention of a specialist, direct interoperability with analysis tools. Moreover, the system is able to operate in different contexts, including medical and non medical domains.

We explain the functioning of our system through Sections II-VI, and then we present, in Section VII, the extensions that we are currently carrying out to obtain a general and flexible system.

II. OVERVIEW ON THE SYSTEM

1) General specifications. The proposed system promotes the collection of health data by the patient himself. The location of the patient at home or on the move is irrelevant. The only constraint is that the patient has to establish an Internet connection at regular intervals (for example, every evening). Two modes of collection are possible: on one hand, a collection by digital questionnaires via smart phones or tablets, on the other hand, automatic collection by using sensors embedded on smart phones or connected watches. All data is transmitted to a central server in a tabular format (compatible with Excel) for storage and subsequent analysis by a software tool, such as SAS (Statistical Analysis Software) [18] or R [19]. It must be possible to carry out analysis relating to a patient or a group of patients.

The creation of the survey questionnaires is carried out by members of the medical staff. A user-friendly interface is therefore available to perform this task. It is important that this interface can offer a good variety of question types.

Each patient responds to one or more survey questionnaires via the tablet or smart phone according to a pre-established timeline. Notifications are generated by the system as soon as a questionnaire is open. Physical activity (number of steps per hour or per day) is captured via an accelerometer installed on the watch. The advantage of the automatic capture of the activity results from the observation that the survey questionnaires do not allow a reliable collection. Experiments have shown that patients systematically overestimate their physical activity. Tracked data is transmitted from the tablet or smart phone to the central server via a certificate-based protocol using a patient specific identifier. The exchanges always take place on the initiative of the mobile devices.

It is the members of the medical staff who manage the patients (and in particular the assignment of an identifier to each patient) via a specific module.

The system must be simple to use so that its acceptability by the users (members of the medical staff on the one hand, patients on the other hand) does not pose any problem.

Mobile devices may be provided by the patient or by the medical service. The patient provides the Internet access device.

2) Architecture. The chosen architecture is simple (Figure 1). It is based on a central server that hosts the data and the questionnaire and the main computer application (called the server application thereafter). This server is installed in a protected intranet. Each patient has a tablet (possibly associated with a watch, or a bracelet, or other sensors) that can exchange data with the server via a secure Internet protocol. It is the mobile application installed on the tablet that initiates all exchanges with the server.

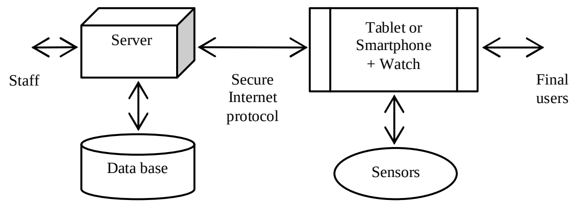

Figure 1. Architecture of the system

3) Main operations. The main operations permitted by the system take place chronologically as indicated below.

① Creation of questionnaires (actor: medical staff in intranet).

② Creation of patients (actor: medical staff in intranet).

③ Assignment of questionnaires to patients (actor: medical staff in intranet).

④ Initialization of the mobile devices (actor: medical staff in intranet): The mobile application is installed on the patient's tablet and watch. The server connection information is initialized on the tablet. The questionnaires are transferred to the tablet.

⑤ Initialization of the tablet connection to the Internet network (actor: patient).

⑥ Response to questionnaires and possible activation of the watch (actor: patient): The data is stored temporarily on the watch and the tablet.

⑦Transmission of data to the server (actor: mobile application in the Internet): The data is transmitted to the central server as soon as the Internet connection is established. Connection and transmission are fully automatic. It is the mobile application that drives the exchanges.

4) Technologies. For the server, we chose a WINDOWS technology associated with MYSQL [15] to manage data storage. The server application is encoded in Java. For mobile devices (tablet or smart phone, watch), we chose an Android technology [16] associated with SQLite [17]. The advantages of Android are two-fold: great variety of mobile devices supported by this system, affordable prices. The mobile application is also encoded in Java. The tablet and watch are interconnected in Bluetooth mode. Data exchanges are carried out by Web services using REST (Representational State Transfer) technology [20].

III. COMPUTER APPLICATION FOR THE SERVER

1) General interface. The general menu (Figure 2) contains tabs for managing medical staff, managing patients, managing survey questionnaires, assigning questionnaires to patients, initializing collection, reporting about a patient or a survey.

2) Model for questionnaires. As soon as a questionnaire is created, a name is assigned to it (this name is used to locate it on the tablet's home page). A questionnaire can be divided into sections. A section may be submitted to the user several times in the form of a series of predetermined deadlines or in the form of a regular repetition over time. A section can include different types of questions: multiple choice questions, cursor questions, grid questions, open-ended questions. These types correspond to those which are most frequently encountered in the medical field. Other types can be added with the same specification approach. The wording of a question is handled in two formats: a long format that is the full text of the question as it is displayed on the tablet screen, and a short format that is used to locate the question in the result table.

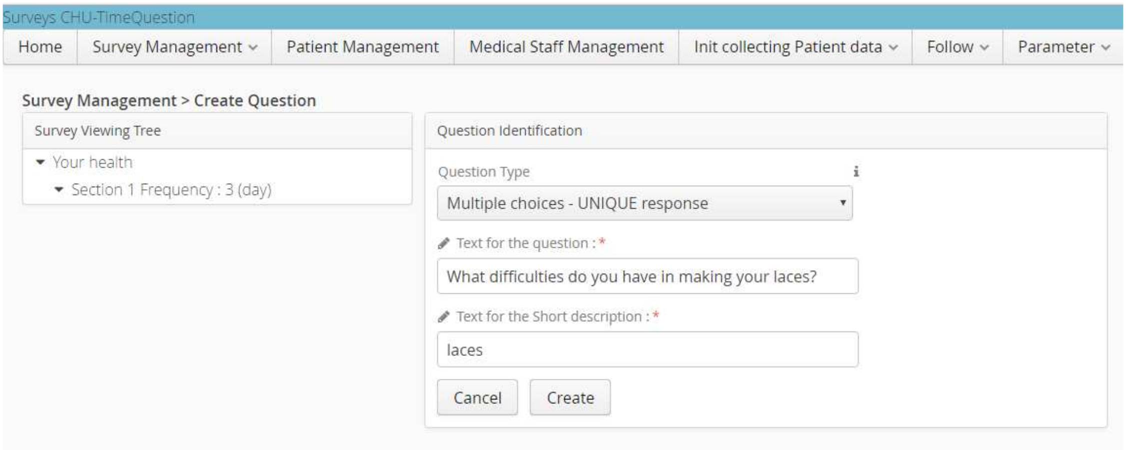

Figure 2. Main menu for the server application

Once the creation of the questionnaire is validated, it switches to the "finalized" state. This status means that the survey is ready to be published to patients.

3) Assignment of questionnaires to patients. Any finalized questionnaire may be associated with patients. The "Assignment Survey" tab allows to associate surveys with patients. It is possible to specify several associations simultaneously.

4) Initialization of the mobiles devices of a patient. The initialization of the mobile devices is obtained by scanning a QR (Quick Response) code generated by the server application. This QR code consists of the patient id, the initialization date, the web service address on the server that must be used for the data exchanges.

5) Results management. We separated the results of the surveys and the results of the sensors in two different classes: *result* and *sensorResul*t. The typing of the collected data is the same for the two classes and respects the following format:

(patient id, variable name, collected value, date)

The variable name corresponds to the short label of a question or to the label of the parameter collected by a sensor.

6) Reporting about a survey or a patient. The results for a survey (all patients combined) or the results for a patient (answers to questions and parameter values coming from sensors) can be downloaded in a same Excel file. An example is given in Figure 3. From this file, we can then make ad-hoc reporting or in-depth analysis by using a tool such as SAS or R. It is interesting also to note that such a file can be seen as the fact table of a warehouse with three main dimensions : patients, variable, time. Approaches proposed for calculating indicators in data warehouses can thus be usefully exploited.

patient id	variable name	value	date
4	Steps per hour	151	12/12/2016 18:01
4	Difficulty : Go down the stairs	average	12/12/2016 18:50
4	Difficulty : Walking flat	low	12/12/2016 18:50
4	Difficulty : Thread the socks	average	12/12/2016 18:50
4	Difficulty : Get out the bed	low	12/12/2016 18:50
4	Steps per hour	41	12/12/2016 19:01
4	Steps per hour	205	12/12/2016 20:00

Figure 3. Reporting about a patient (excerpt)

IV. COMPUTER APPLICATION FOR THE MOBILE DEVICES

1) Answering a questionnaire. The names of the different surveys associated with the patient are displayed in different bannners on the tablet home page (Figure 4). These names are those that were specified when creating questionnaires. Banners with a gray background correspond to surveys that are not due at the time of the consultation and are therefore inaccessible. The patient can answer any survey accessible by clicking on the corresponding banner. The accessibility of a survey is determined in accordance with the frequency or timelines specified at the time of its creation.

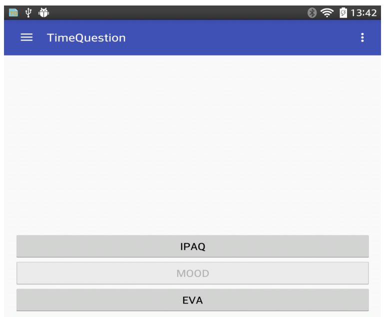

Figure 4. Accessing surveys from the tablet home page

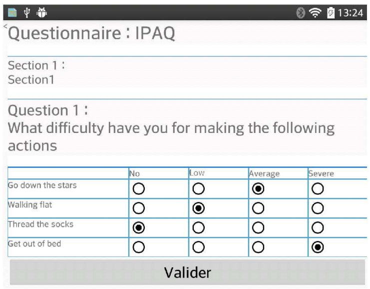

Figure 5. Answering a question on the tablet (grid question)

Figure 5 shows the presentation of a grid question on the tablet screen.

2) Step counting. The accelerometer of the watch and the Google Step component are used to count the steps. A higher layer was developed to aggregate the count over each 60 minutes interval. It is the value of this aggregate that is transmitted to the server via the tablet.

Other medical sensors are embedded on Android supports and can be installed on our system. They permit to capture parameters such as Temperature, Blood Pressure, Pulse, and Heart Rate.

3) *Data exchanges*. The sending and receiving of data between the watch and the tablet, on one hand, between the tablet and the server, on the other hand, is done automatically without user action. A short message is displayed at the bottom of the page for a few seconds to signal the shipment. The sequence diagram below (Figure 6) describes the principle of exchanges between the mobile devices and the server. The data of the watch is first stored in its internal memory and then transmitted on its initiative to the tablet. The data generated at the tablet level (ie the answers to the questions) and those recovered from the watch are stored in its internal memory and then transmitted to the server at its own initiative. Initialization of the exchange by the transmitter (watch or tablet) occurs every 30 minutes if the network is available (Bluetooth for communication to tablet, Internet for communication to server). The transmitter keeps the data until the receiver has returned an acknowledgment of receipt. If this acknowledgment fails within 30 minutes, the sender attempts a new sending. When the acknowledgment is received, the sender removes the data from its internal base.

4) *Deployment on the playstore*. Our mobile application, called TimeQuestion, was deployed on the playstore to simplify the propagation of updates and initialization. TimeQuestion includes the codes to be installed on the tablet and the watch. The installation of TimeQuestion is done automatically on the tablet and also on the watch, if a watch is connected. The application is optimized for a 7-inch tablet, but it can also be installed on a smart phone.

V. EXPERIMENTS

The system is now fully operational and the server is permanently active. Various experiments have been carried out with members of the Physical Medicine and Rehabilitation Department of the Clermont-Ferrand Hospital (France) and patients of this department in order to evaluate the acceptability of the system.

First, multiple demonstrations were carried out by the authors. Once the mobile devices have been initialized, it is no longer necessary to worry about them. If they become inactive following a discharge of their batteries, simply recharge the batteries and restart them.

Several patients were asked to test the system when they were in the PMR (Physical Medicine and Rehabilitation) Department.

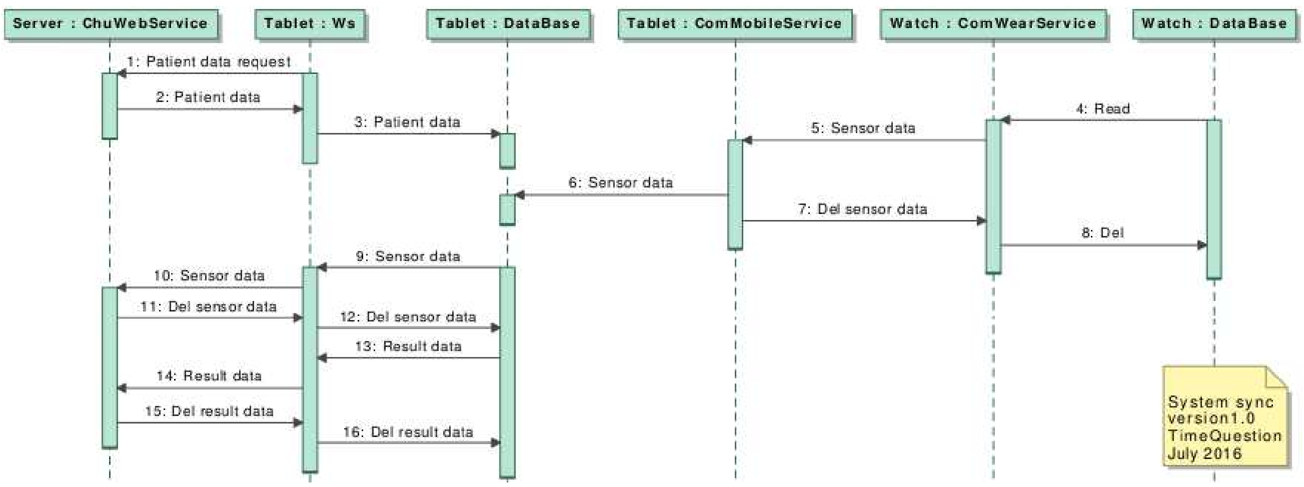

Figure 6. Data exchanges between the mobile devices and the server

 ISBN: 978-1-61208-568-5

One patient was asked to use the system for two weeks. The PMR (Physical Medicine and Rehabilitation) Department has put at her disposal a tablet of 7 inches as well as an android watch. A survey with 16 questions relating to movement difficulties was assigned to her. A response was requested every two days. This patient was asked to walk 30 to 60 minutes a day. We were able so to verify that the data went back to the server on a regular basis. Figure 3 illustrates an excerpt of the collected data during this experiment.

All these actors found that the system was very convenient to use and very useful.

We did not have the opportunity to address the acceptability of the system by patients with disabilities, such as those with hand tremors or those with visual deficiencies. We are confident about the efficiency of our system for those with hand tremors because tablets with large screen sizes can be handled by the system. We have not explored solutions for patients with visual deficiencies. For this type of handicap, specific solutions need to be studied.

The MPR department studies the effects of hydrotherapy for its osteoarthritis patients. It planned to use the system with about ten patients for the next treatment periods.

VI. TOWARDS A MORE GENERAL AND FLEXIBLE SYSTEM

We explain in this section the extensions we have undertaken to make the system more general and more flexible.

1) *Separate management of questions*. In the current version, each question is linked to a questionnaire. Experiments have shown that the same question may appear in different questionnaires. It is therefore a question of being able to specify the questions separately, then to assemble them to form the questionnaires.

2) *Semantic standardization of labels*. In the current version, the wording of the questions and parameters captured is left to the free choice of the medical staff. It is then difficult to integrate the data coming from different services (remember that these labels serve as semantic reference for the collected data). Yet, this integration would be interesting to deal with more massive data coming from different horizons. To facilitate this integration, we propose to constrain the choice of labels through a shared ontology of the domain. We are adapting the specification of long and short labels in order to impose the choice of a term in an ontology.

3) *Sensor assisted installation*. In the current version, each sensor is associated with a specific software component which, collects the raw data transmitted, ensures their filtering and aggregates this data over the relevant period before transmitting them to the tablet. In addition, the label of the sensed parameter is hard-coded in the component. The extension consists of decomposing this component into two parts: a part that remains specific to each sensor (this part collects the raw data and performs the filtering), a part that ensures the naming of the associated parameter and the calculation of the aggregate. The main menu of the server application is redesigned to allow the choice of the sensor, the choice of the associated label, the choice of the type of aggregation to be performed (sum, average, etc.), the aggregation time interval. The second part of the component can be then automatically generated. The installation of a sensor can be so specified directly by a user manager without requiring the intervention of a developer.

VII. CONCLUSION

The initial objective of this work was to design and develop a system for the self-collection of medical data. The data come from the responses to questionnaires transmitted via the tablet, on the one hand, and the parameters collected by the watch via sensors, on the other hand. The questionnaires are defined by the members of the medical staff through a convivial interface and their structures are stored on the server in a relational database. They are then loaded onto the tablet during an initialization procedure. The application allowing the reading of the values of the sensors is also automatically installed on the watch during this initialization. All data collected from questionnaires and sensors are stored in a unified tabular format to facilitate their recovery by a spreadsheet in order to activate various statistical analyzes or data mining treatments.

Our system has also other main advantages. It allows full mobility of the patient. Its management can be handled by a member of the medical staff without the intervention of a specialist. Its operation is automatic as soon as the initialization of the mobile devices has been carried out.

We have conducted experiments which have shown that the system is well accepted by the patients.

It is interesting to note that it is possible to incorporate other types of data into our warehouse. For example, in the medical domain, the fact table could be used to store medical analysis results or imaging reports. Technically, we need to study the interconnection of our system with the other systems used by physicians.

The data collected, in particular via the sensors, can quickly become bulky, and one can wonder about the suitability of such a system for handling big data. The main problem is the server's storage capacity. A relational table under Windows NTFS (New Technology File System) has a maximum capacity of 2GB. We can evaluate the length of a line in the *result* table or the *sensorResult* table to 200 Bytes. Suppose the server is used by a medical department to track 100 patients. It is thus possible to store 200,000 answers to questions for a single patient and to store the number of steps per 60 minutes over 25 years for a patient. It is very comfortable. But if we want to integrate data coming from several departments into hospitals across the country, this capacity may become insufficient and other storage technologies should be considered. Today, there are technologies for big data that remain compatible with our architectural choices and that do not put into question our software.

This system was initially defined on the basis of wishes expressed by a hospital department. But it has been designed and developed in a generic way and can be used to collect any kind of data from an individual or a natural or artificial entity. First, it can be used for monitoring individuals in various situations: athletes in training, workers in the performance of certain tasks, etc. But, it can also be used for monitoring any type of non human entity. The sensors are then installed on the entity and the questionnaires are activated by a human observer of this entity. For example, sensors can be used to monitor vegetal growth and collect immediate environmental conditions (e.g., moisture and temperature for air and soil). Questionnaires can collect more environmental information (e.g., nature and evolution of the surrounding plantations), useful for explaining the vegetal growth.

This system is currently being extended to make it more general and flexible. These extensions mainly concern three directions: the separate specification of the questions, the semantic standardization of the labels identifying the collected data, the assisted installation of a sensor.

REFERENCES

[1] C. Schmitz, "Lime Survey: the free and open source survey software tool", Available on: http://www. limesurvey. org/, Accessed on June, 12, 2017.

[2] D. H. Granello and J. E. Wheaton, "Online data collection: Strategies for research," Journal of Counseling & Development, no 82(4), pp. 387-393, 2004.

[3] K. B. Wright, "Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services," Journal of Computer-Mediated Communication, no 10(3), 2005.

[4] A.Tellegen and N. G. Waller, "Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire," The SAGE handbook of personality theory and assessment, no 2, pp. 261-292, 2008.

[5] D. Y. Leung and D. Kember, "Comparability of data gathered from evaluation questionnaires on paper and through the Internet," Research in Higher Education, no 46(5), pp. 571-591, 2005.

[6] M. N. van Poppel, M. J. Chinapaw, L. B Mokkink, W. Van Mechelen and C. B. Terwee, "Physical activity questionnaires for adults. Sports medicine," no 40(7), pp. 565-600, 2010.

[7] K. Kroenke, R. L. Spitzer, J. B., Williams and B. Löwe, "The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review," General hospital psychiatry, no 32(4), pp. 345-359, 2010.

[8] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), no 40(1), pp. 1-12, 2010.

[9] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," Computer Networks, no 54(15), pp. 2688-2710, 2010.

[10] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola and I. Korhonen, "Activity classification using realistic data from wearable sensors," IEEE Transactions on information technology in biomedicine, no 10(1), pp. 119-128, 2006.

[11] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," IEEE Communications Surveys and Tutorials, no 15(3), pp. 1192-1209, 2013.

[12] A. Gaggioli, G. Pioggia, G. Tartarisco, G. Baldus, D. Corda, P. Cipresso and G. Riva, "A mobile data collection platform for mental health research," Personal and Ubiquitous Computing, no 17(2), pp. 241-251, 2013.

[13] D.R. Bassett Jr, "Validity and reliability issues in objective monitoring of physical activity," Research quarterly for exercise and sport, no 71, pp. 30-36, 2000.

[14] R. Rubin and M. Peyrot, "Treatment satisfaction and quality of life for an integrated continuous glucose monitoring/insulin pump system compared to self-monitoring plus an insulin pump," Journal of diabetes science and technology, no 3(6), pp. 1402-1410, 2009.

[15] J. Greenspan and B. Bulger, "MySQL/PHP database applications," John Wiley & Sons, Inc., 2001.

[16] R. R. Lombardo, J. Mednieks and B. Meike, "Android application development: Programming with the Google SDK," O'Reilly Media, Inc., 2009.

[17] M. Owen and G. Allen, "SQLite," Apress LP, 2010.

[18] R. C. Littell, "Sas," John Wiley & Sons, Ltd, 2006.

[19] R, "The R Project for Statistical Computing," https://www.r-project.org/, Accessed on June, 12, 2017.

[20] C. Riva and M. Laitkorpi., "Designing web-based mobile services with REST," Service-Oriented Computing ICSOC 2007 Workshops, Springer Berlin/Heidelberg, 2009.