Prototype System for Improving Manually Collected Data Quality

JARI SOINI, PEKKA SILLBERG and PETRI RANTANEN, Tampere University of Technology – Pori

Even nowadays, a great deal of measurement data is collected and also saved manually. In this kind of situation, there are phases when human error can easily occur and also when interpreting the typed collected measurement data could be difficult. This research aimed to discover resources for improving the quality of measurement data as well as better and more illustrative tracking of usage information in real time. The objective was both quality improvement of a specific measurement data collection process as well as the elimination of human error. This paper describes one reliable solution for this purpose, which improves the quality and also the visual presentation of manually collected data. The paper presents elements of the system developed for this aim and also the technology deployed along with its operational principles.

Categories and Subject Descriptors: H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*; H.4.0 [**Information Systems Applications**] General

Additional Key Words and Phrases: Data quality, measurement process quality, data visualization, software applications

1. INTRODUCTION

The starting point of the research was to map out areas and activities of the public sector in which *savings* could be achieved by controlling, optimizing and intensifying operations. This research is a part of the ongoing two-year (2013-2014) Kiiaudata (Kiinteistöjärjestelmien datan älykäs analysointi – smart analysis of property systems data) project funded by Tekes [2014], where one of the main aims was to study potential new technologies for managing and controlling conditions in buildings in a smart way. In collaboration with the City of Pori, a survey was made about the points where measurement data is collected and also how said data is utilized. As the result of this mapping, it was decided to focus on the upgrading of measurement data collection and the new swimming pool was chosen as the research subject, as it is the city's most expensive individual building in terms of energy consumption.

The idea was that the maintenance staff would continue checking the physical measuring devices to ensure their conditions, but the collected data would be recorded with the developed system in contrast to the fully manual record keeping used in the past (i.e. pen and paper). The measurements produce information that can be used, for example, in consumption and condition tracking. For instance, analyses of alteration in energy consumption can be made by means of inclusive measurement and usage tracking based on it. Electricity, heat and water are examples for different measured energy currents. In many cases, the aforementioned currents can be tracked and anomalous situations can be reported automatically using modern computer controlled systems, but there still remain situations where manual work is required, especially when dealing with legacy systems.

Author's address: J. Soini, P. Sillberg and P. Rantanen, Department of Software Engineering, Tampere University of Technology – Pori, P.O.Box 300, FIN-28101 Pori, Finland; email: {jari.o.soini, pekka.sillberg, petri.rantanen}@tut.fi

In: Z. Budimac, T. Galinac Grbac (eds.): Proceedings of the 3rd Workshop on Software Quality, Analysis, Monitoring, Improvement, and Applications (SQAMIA), Lovran, Croatia, 19.-22.9.2014, published at http://ceur-ws.org.

There are several studies related to building automation systems and automatic sensor data collection, for example Cheng and Shen [2011] introduced wireless sensor networks based on embedded Linux. Nainwal et al. [2011] studied on remote surveillance and monitoring system utilizing wireless sensor networks, Vujović and Maksimović [2014] focused on utilizing Raspberry Pi as a building block of wireless sensor node, and Toshniwal and Conrad [2010] introduced a web-based sensor monitoring system on a Linux-based single board computer platform. However our focus was on systems where automatic sensors cannot be fully utilized. The work presented in this paper utilizes the findings of Soini et al. [2013], in which mobile devices, Global Positioning System (GPS) technology and route optimizations were combined in a real-time tracking service for delivery of goods.

The owners of the property chosen as the research subject – the new public indoor swimming pool of the City of Pori – were particularly interested in, for example, identifying development targets related to energy consumption measurement, development of the measurement process, early discovery of possible issues, and evaluation of the impacts of changes. For this research, a manually used digital data collection system has been developed as a collaboration project between Tampere University of Technology (TUT) and the City of Pori. The system developed facilitates the maintenance staff's work in registering and recording the measurement information as well as real-time tracking of usage information and perception of possible anomalous consumption situations.

2. PROBLEMS IN QUALITY OF MANUALLY COLLECTED DATA

Erroneous values are common when collecting and typing up data by hand, especially for long numeric values. Errors can also be very hard to detect, and it is difficult to know if the erroneous value was caused by an error with a meter or a correct value was simply mistyped by the person reading the meter. This was the problem observed and the starting point of this study. The assumption was that typing errors can be detected by software.

In some cases, it is not financially viable to replace measuring devices: many devices available today can be networked and contain automatic error detection or monitoring software, but this is not true for all devices, especially when taking into consideration many legacy devices. If these devices are seldom used or replacing them would be expensive, alternative approaches are required.

There are still many measuring devices that need to be checked periodically by a user. In practice this may require writing down the values by hand. In many places it is still common to use the basic pen-paper-and-Excel approach, in which the measurements are checked manually, written down and later inputted using datasheet software such as Microsoft Excel. The system presented here enables the pen-and-paper phase to be skipped. Using a management interface, reports of the values can be created and saved in various formats (such as .pdf or .xls). The paper describes simple client software, which uses Near Field Communication (NFC) [ISO 2013] tags to detect a measurement device called an "object" in the context of this paper. In the scope of this paper, an object means a monitored physical device (e.g. water meter).

3. SOLUTION – PROTOTYPE SYSTEM FOR COLLECTION OF CONSUMPTION DATA

The main idea behind the prototype system is to combine a typical web service, a mobile device with networking capability and a way to identify the object to implement a data gathering and reporting service. QR-Codes and Radio Frequency IDentification (RFID) [ISO 2008; Finkenzeller 2010] contactless proximity cards were the main candidates for identification purposes. RFID cards were chosen over QR-Codes as they should be more reliable to recognize in dimly lighted environments. It is also more convenient to touch the card instead of taking a photo of QR-Code when the space is limited.

Not all RFID cards, or tags, are alike as they vary on parameters such as operating frequency, data speed, distance of reading, power supply (passive, active, battery-assisted passive), and price. The

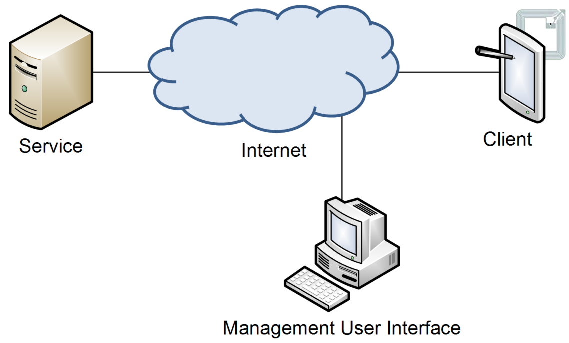

Fig. 1. System overview.

choice of parameters depends on the use case [Nummela 2010]. Typically, a low operating frequency correlates with low data speed and reading distance. An active power supply increases the price of the tag but enables the tag to operate without the support of a tag reader. We chose to use Near Field Communication (NFC) compatible tags as they are:

—relatively inexpensive

—they receive all the required power from the reader which reduces the need for maintenance

—the reading distance was not a crucial part of the system

—NFC capable smart phones and tablets are becoming more common.

For the purpose of this application, we are only interested in the unique ID which can be read from every tag. In our system this ID – i.e. a tag – is bound to an object. The user only has to touch the tag and the client software retrieves the correct data. Every object can be configured with various details:

—a common name (e.g. Water consumption)

—names of related gauges (e.g. Main water meter)

—the unit of the gauge (e.g. Cubic meters)

—warning limits for expected minimum and maximum daily increase (e.g. we expect that the gauge reading could increase by 50 to 100 units per day).

Figure 1 shows an overview of the system. The *Service* is available over the Internet where both *Management User Interface* and *Client* application can be connected. The service uses JavaScript Object Notation (JSON) to transmit data objects and it has two Representational state transfer (REST) interfaces, one for getting the gauge data and the other for posting the gauge data. It also supports user access control, but this feature is not currently used in the pilot phase of the system. The management user interface is a JavaScript-based web page accessible with a web browser. There the system administrator can configure a particular object and interpret the results sent by the client. For example, the results can be viewed as raw data or plotted as a chart. The *Client*, in Figure 1, is the main component that the end user is using. It is used to interact with tags, collect the data, and perform small scale on-site analysis of the data. The client application in our case is programmed for Android devices.

The prototype system is currently being tested in at the new swimming pool in City of Pori. There are three different gauges (water, electricity and central heating) which are being monitored. There are a couple members of the maintenance staff who operate the client device just to collect the gauge readings, and one person to oversee the changes in the collected data. All workers are operating the

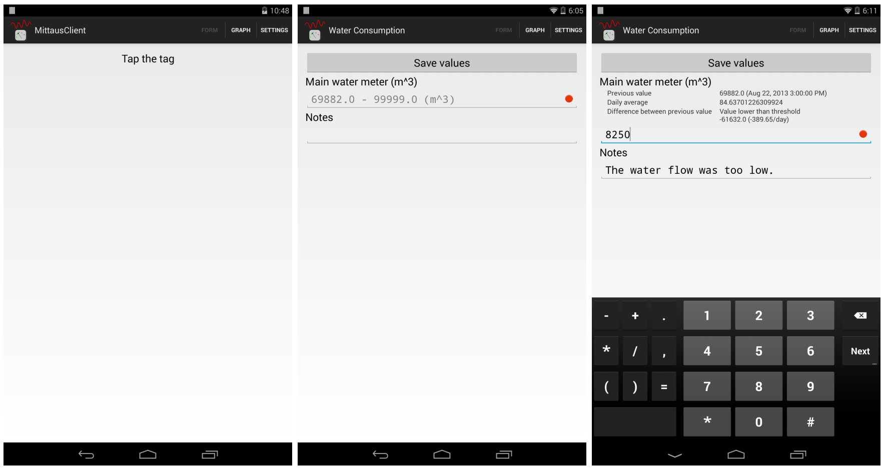

Fig. 2. Application screenshots, from left to right: initial view before a tag has been read, form view after the tag has been read, and finally, threshold value is below the defined limit.

client thru one shared device. So far the response from the staff has been enthusiastic about the data collection system, particularly of the ability to see the approximate costs of the facility immediately.

3.1 Information Collection

The client device and the NFC tag play an important role in collecting the meter readings. The information collection consists of three phases:

(1) identifying the object
(2) inputting the data
(3) saving the data.

Each of the three phases is explained in more detail in the following sub-sections.

3.1.1 *Identifying the Object.* The first step is to identify the object by touching the tag attached to the object. The tag will be automatically detected by the device. The tag detection is based on the unique ID found on every tag. In the current implementation these IDs can be mapped to objects using the service's management interface. This mapping is used by the client to detect which object is the current target and to show the correct object-dependent input fields. It could also be possible to extend the client software to enable mapping new tags for objects, which would make installing the overall system easier. This way the system installation could use bulk tags, which would be mapped to objects on the spot by the person performing the installation procedure. Whether mapping the tags on the device is required depends on the use case, and in our current scenario it was not a necessary feature mainly because of the relatively small amount of objects and tags. Also, as the main use of the client device is to gather information, it might be better to keep the software simpler to use by limiting the functionality available (see Figure 2).

The mapping information and the input field details can be synchronized with the service at any time, but in general, synchronization is performed only when specially requested. There are two reasons for this: firstly, the mapping and input field details change very rarely, making continuous synchronization a waste of network bandwidth; and secondly, in some cases the objects may be located in places with poor or non-existent network connectivity, making live synchronization difficult or even impossible. The basic view before any tags have been detected is shown in Figure 2 (left), and the view after a tag has been selected is shown in Figure 2 (center). In the example case a very simple object is illustrated containing only two fields; a numerical input field for the *Main water meter*, which accepts values ranging from 69882 (the previous input value) to 99999, and a text input field for *Notes*.

3.1.2 *Inputting the Data.* Figure 2 shows the views of a detected object. The view in the center shows the basic view and the view on the right show the extended view. When the user taps any of the fields, additional information related to that specific field is shown: the previously given value with the timestamp of the input date, the daily average, and the difference of the currently typed value (if any) in relation to the previously given value. The purpose of the extended information is to give a quick glimpse of previous data, which can be used to detect possible errors in the readings and give the person using the device an idea of the possible values. In the example case (Figure 2, right), the red dot on the right hand side of the input field shows that a bad value has been given, and the user has typed a descriptive comment on the matter in the *Notes* section ("The water flow was too low"). Figure 3 (left side) shows the same case with the properly inputted value.

The value ranges used to detect and show warning situations are configured on the management web interface of the service. The ranges are numerical thresholds, which have either been calculated based on earlier data (e.g. it may be known how much water is used on average on a daily basis), or they may be based on physical limits (e.g. water consumption cannot be negative). The ranges can be simple minimum and maximum values, which should not be exceeded (e.g. voltage should stay between 10 and 15 volts) or cumulative limits (e.g. water consumption should not exceed ten cubic meters per day). The minimum and maximum values do not need previous values for accurate calculation of the warning threshold. In the case of cumulative limits, at least one single previous value is required. The previous values can be provided by the service when synchronizing the tag mappings and input fields or they can be results from previous use of the software. The warnings are meant to help the person typing the input values, and they are only "soft limits", i.e. they can be overridden if required. For example, it may be possible that a meter is giving an erroneous reading or for some reason much higher consumption is occurring. In this case it may be required to input a value that is outside the previously designated range. Inputting a value outside the range requires a confirmation from the user, and it will automatically be detected by the system and will pop up as an erroneous value on the management interface. It is also possible to generate an automatic notification, for example an email or SMS alert to be sent when an erroneous value is detected by the service, but in practice the notification will not be sent immediately if the data inputting process is performed in a location without network connectivity.

3.1.3 *Saving the Data.* After the user has inputted the desired data, the *Save values* button can be used to submit the results. The submit process may not necessarily start immediately. The values are stored locally on the device and can be viewed at any time, but when the actual result submission happens depends on the availability of the service. The client contains a background service which will periodically try to submit any unsent information. In our use case, the measuring devices themselves are located in an area of poor connectivity, but the users' workplace contains areas where the results can be submitted. The users generally carry the client device with them, thus allowing the automatic submission of the results when a network connection has been established. If instantaneous submission is required, other approaches should be considered such as providing wireless access by using a wireless router. The effects of periodic submit retries on battery life may vary. On one hand, turning the wireless radio off, and turning it on only when required may improve the battery life of the device. On the other hand, if the availability of the network connectivity is unknown, it may be difficult to establish the connection at timed intervals. In practice, many tablet and smart phone devices can

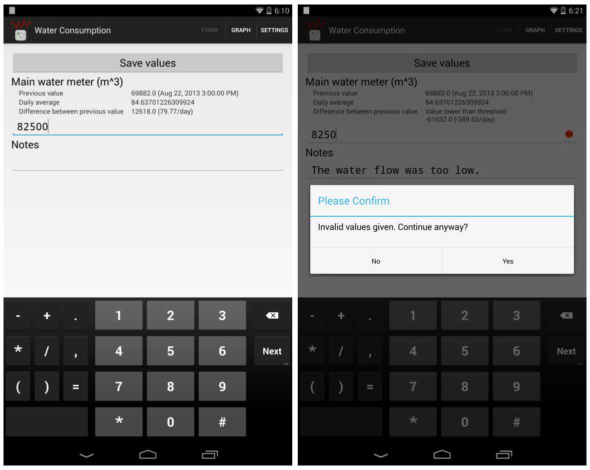

Fig. 3. Application screenshots: the left figure is the form view with data given by the user, the right figure asks for user confirmation before saving the data.

sustain a battery life of a whole day using the default power saving settings, thus only requiring the device to be charged when not needed, for example, outside working hours.

If the inputted values contain erroneous out-of-range values, a confirmation of values is required before the data can be saved and sent. The confirmation dialog is illustrated in Figure 3 (right).

3.2 Viewing the Results

The system allows the user to examine the collected data quickly on the client application and more thoroughly using the management user interface. Figure 4 illustrates the general idea of the different views:

—simple chart view of the client application on the left

—more complete analysis chart of the management user interface on the right.

The rationale for limiting client application features is to keep it as simple as possible and therefore to reduce the maintenance required for the application. It also helps to keep the device small enough for carrying around and for entering data. Also, the employee typing in the data might be more interested in seeing if the figures show any unexpected highs or lows, so he/she can react to the situation more quickly.

Both charts in Figure 4 contain the same data (consumption of water; x-axis time; y-axis consumption in cubic meters), but the view on the client application (Figure 4, left) is panned and zoomed in to show data between June 2013 and August 2013. The browser view (Figure 4, right) displays all of the data beginning from January 2012 and ending in May 2014. The upper chart shows the actual data and the lower chart illustrates the calculated daily average consumption. Between the charts there is a section with statistical information about the consumption. It shows the meter reading, date of the reading, and also approximated daily, weekly, and monthly costs in euros (by using a predefined average price per unit). The statistical information follows the pointer of the mouse so it is possible to see the same data from any point of the chart.

A surge in water consumption can be seen during July 2013 with consumption peaking at 400 cubic meters per day. This kind of information can be helpful for the maintenance team as it could be a sign of a leakage somewhere in the system. Fortunately, the peak was due to a scheduled maintenance of

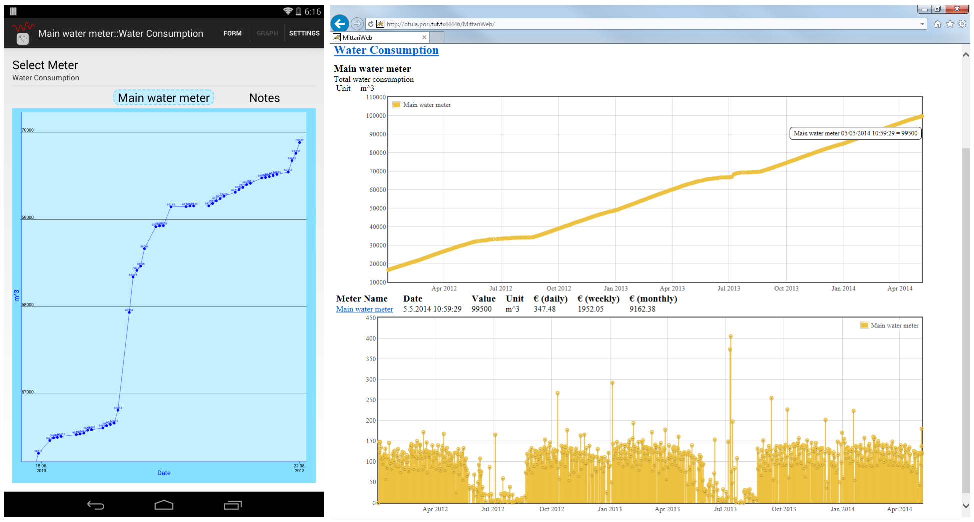

Fig. 4. Chart views as seen on mobile application (left) and on web browser (right).

the swimming pools. There are also many small consumption peaks and lows on the lower chart of the browser view. This occurred because the data was imported from the handwritten notes without exact time information. The collection time of the imported data is simply set at 12 noon, so it will cause fluctuation if the meter was actually read in the morning or evening. In the future as the data is collected directly to the system, the exact reading time can be stored, which will eliminate the fluctuation caused by unknown meter reading times.

The data shown in Figure 4 has been imported into the system from the actual water consumption data collected from the new public swimming pool located in the City of Pori. The facility has also been recording the consumption of central heating and the consumption of electrical energy. As the data comes from an actual facility, we had the opportunity to reflect on the consumption in terms of what had really happened. The data can be broken down into the following sections (see Figure 4, right side):

(1) January 2012 – June 2012, (winter & spring season, average consumption)
(2) June 2012 – August 2012 (summer maintenance, low consumption)
(3) August 2012 – June 2013 (fall, winter & spring season, average consumption)
(4) June 2013 – August 2013 (summer maintenance, from low to high consumption)
—Contains a surge of water consumption due to pools being emptied, overhauled, and then refilled.
(5) August 2013 – May 2014 (fall, winter & spring season, average consumption)

The data has been recorded by pen-and-paper, but is now being stored directly on the electronic database by using the system described in this paper. In fact, there are a lot of other digitally monitored and configurable parameters in the new swimming pool facility, but these three gauges (water consumption, central heating consumption and electrical energy consumption) are the only meters that still require old-fashioned manual reading.

4. DISCUSSION

The efficiency of the system greatly depends on the defined value ranges. If it is not possible to define clear ranges or the ranges remain vague, the possibility of error increases, and in this case the software works only as a pen-paper-and-Excel replacement. In practice, based on user feedback, the most common source of error was grossly mistyped numbers, caused by lengthy numeric values (e.g. when writing down values it is easy to mix up 154763 and 157463, an error that can easily be detected by the software).

The software is more suitable for use cases where the meters are not read very often, but *do need* to be read manually periodically. If the meters need to be read continuously, for example several times a day, it may be more advisable to invest in meters with an automatic monitoring and warning system (if possible). On the other hand, if the meters are hardly ever checked, the basic pen-paper-and-Excel approach may be more feasible, and the resources required for setting up the system can be saved.

Then why not change the remaining analog meters? The comments from the facility's maintenance workers were that if they routinely read the meters every day, they can simultaneously monitor the condition of the nearby equipment and perform preventative maintenance if needed. Thus they can complete several tasks at once. It also helps to get a better grasp of the facility as a whole as they can see how much power or water is consumed daily.

5. SUMMARY

The paper presents a system for improving the quality of manually collected data. In many cases, especially in the public sector, there are many different points where manually measurement data collection is still practised. These situations usually relate to the monitoring of the operations of some physical devices, such as energy-related consumption measurement. The system introduced assists maintenance staff and also supports managers who are responsible for ensuring the correct operation of the devices. This system is one step towards more reliable and thus better quality measurement data, and it also improves the visual presentation of collected data for analysis. During the ongoing study, the system features will be extended and adapted for the purpose of monitoring patient rooms in the public sector health care environment.

REFERENCES

Xiaohui Cheng and Fanfan Shen. 2011. Design of the wireless sensor network communication terminal based on embedded Linux. *2011 IEEE 2nd International Conference on Software Engineering and Service Science* (July 2011), 598–601.

Klaus Finkenzeller. 2010. *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication* (3rd ed.). Wiley.

ISO. 2008. *ISO/IEC 14443, Identification cards – Contactless integrated circuit cards – Proximity cards*.

ISO. 2013. *ISO/IEC 18092:2013, Information technology – Telecommunications and information exchange between systems – Near Field Communication – Interface and Protocol (NFCIP-1)*.

V Nainwal, P J Pramod, and S V Srikanth. 2011. Design and implementation of a remote surveillance and monitoring system using Wireless Sensor Networks. In *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, Vol. 5. 186–189.

Jussi Nummela. 2010. *Studies towards Utilizing Passive UHF RFID Technology in Paper Reel Supply Chains*. Doctoral dissertation. Tampere University of Technology.

Jari Soini, Timo Widbom, Jari Leppäniemi, Petri Rantanen, and Pekka Sillberg. 2013. Pilot system for transport confirmation with location awareness. In *Symposium GIS Ostrava 2013 - Geoinformatics for City Transformation*. Ostrava.

Tekes. 2014. Finnish Funding Agency for Technology and Innovation. (2014). http://www.tekes.fi/eng

Kailash Toshniwal and James M. Conrad. 2010. A web-based sensor monitoring system on a Linux-based single board computer platform. *Proceedings of the IEEE SoutheastCon 2010 (SoutheastCon)* (March 2010), 371–374.

Vladimir Vujović and Mirjana Maksimović. 2014. Raspberry Pi as a Wireless Sensor Node : Performances and Constraints. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*. Opatija, 1247–1252.