Usability Assessment of the Visualization-Driven Approaches to the HVAC Data Exploration[*]

Evgenia Novikova[1][0000-0003-2923-4954], Polina Belimova[2][0000-0001-8581-4924],
Alena Dzhumagulova[2][0000-0003-2389-2729], Mikhail Bestuzhev[3][0000-0003-1902-4208],
Yulia Bezbakh[2][0000-0002-7396-675X], Aleksandr Volosiuk[3,2][0000-0001-897-9828],
Andrey Balkanskii[2][0000-0003-2663-0599], and Alexei Lavrov[2][0000-0002-5315-8915]

[1]SPIIRAS, Saint-Petersburg, Russian Federation
novikova.evgenia123@gmail.com

[2]ITMO University, Saint-Petersburg, Russian Federationy
belimova.polina95@gmail.com, aledjuna@gmail.com,
jul-ka5@mail.ru, abalkanskij@gmail.com, lavrov@corp.ifmo.ru

[3]Saint Petersburg Electrotechnical University "LETI", Saint-Petersburg, Russian Federation
bestugev94.gmail.com, aavolosiuk@gmail.com

Abstract. Application of the Internet-connected operational devices in the heating, ventilation and conditioning (HVAC) systems has extended the cyber-attack surface by introducing different malicious scenarios. The analysis of the HVAC data may provide insight on typical patterns of the system operations. Implementation of the thoroughly elaborated visualization models may significantly increase the efficiency of the suspicious activity identification in the HVAC systems. In the paper we present the results of the laboratory usability testing of three visualization models used to analyze HVAC data – matrix-based visualization technique, non-linear multidimensional visualization technique RadViz and timeline chart. Matrix-based visualization and RadViz visualization are often used in anomaly detection process, while timeline charts are a traditional way to present operational HVAC data. We describe the experiment design and discuss the results obtained. The usability testing revealed advantages and limitations of these visualization techniques in behavior pattern and anomaly identification tasks. The results can further serve as guidelines for task-dependent selection of a visualization technique.

Keywords: Usability assessment, RadViz, Matrix-based visualization, Time line, Laboratory study, Analysis process, Interface design, Data visualization.

1   Introduction

The IoT technology allows constructing resource-efficient and sustainable buildings that provide comfortable and convenient living environment and appropriate conditions for equipment functioning. The smart heating, ventilation and conditioning (HVAC) systems system could significantly reduce energy, water and other resources consumption by adopting the system functioning to demands and life rhythms of the building inhabitants. The analysis of the HVAC data may provide insight on typical patterns of the system operations. It could be used for setting up the system parameters as well as detecting different anomalies in the system functioning. Anomalies could be a sign of physical degradation of the equipment or fraudulent activity. It has been shown, that the usage of smart HVAC devices for control and analytics purposes provides opportunities to interfere with the physical safety of people and things [1].

The visualization-driven approaches could significantly increase the efficiency of the exploratory analysis of the HVAC data. They could be used to reveal typical HVAC functioning patterns or malicious scenarios, to describe distinctive attack features when no prior information is available.

In survey on design criteria for visualization of energy consumption the authors highlighted the need to use different effective visualization techniques supporting informed decisions [2]. They also defined the "understandability" of visual representation as a mandatory requirement to the data graphical design. The ease of comprehension of information displayed in the visualization by the analyst defines the efficiency of applying visual analytics techniques for pattern and anomaly detection in HVAC data. However, in many cases the existing visual analytics approaches proposed to analyze the HVAC or energy consumption data [3-7] did not undergo any usability assessment focusing both on their efficiency in solving tasks and easiness of their comprehension.

The paper presents the results of the laboratory usability testing of three different visual models proposed to detect functioning patterns and anomalies in HVAC data [7, 8]. These models are RadViz which is a pixel-based visualization that is based on multidimensional data projection, a matrix-based visualization and line charts. In [8] authors combined these three techniques to explore HVAC data, they proposed to use RadViz to detect daily patterns of the HVAC functioning and reveal suspicious deviations in the state of the HVAC system. Matrix-based visualization is used to detect time intervals with anomalous functioning of the system, while line charts provide detailed information. The goal of the laboratory study is to assess how the participants understand these visual models and how they use them during the analysis process.

Specifically, the contribution of the paper to the field of information visualization is the usability assessment of the application of the multidimensional data projection and visualization technique and matrix based visualization to the analysis of the HVAC data.

The rest of the paper is organized as follows. Section 2 discusses the related work on visual analytics techniques for HVAC, energy consumption data analysis. In section 3 we briefly describe the analyzed visualization models. Section 4 presents a

laboratory study scenario, discusses the obtained results and defines the directions of the future research. Conclusions sum up our contributions.

2   Related Work

There is not much research effort devoted to the designing visual analytics techniques for efficient smart building management and investigation of anomalous deviations in its functioning. The majority of the existing techniques focus on the problem of the energy consumption. The most commonly used visualization techniques to solve this task are standard 2-dimensional visualization techniques such as line charts, bar charts and maps [2, 3].

Palm & Ellegård [4] developed visualization-driven approach to the analysis of the energy consumption in order to construct typical behavior patterns of the energy consumers of Sweden. The activities are represented as a sequence of colored bars, where the color encodes the type of the entity's activity. The consumed energy is displayed using traditional bar chart linked to the sequence of activities by one time scale.

Abdelalim et al. [5] propose to visualize energy and mass flows on building level and HVAC system using Sankey diagram. The sensor data are converted into estimated energy flows for each HVAC component, the color of the flows indicate whether it is normal or abnormal. The authors showed that compared to pie charts and other conventional graphs, the Sankey diagrams allows estimation not only the proportion but the direction of the energy flow, therefore they can be useful in evaluating the setting points of the HVAC equipment and investigating the causes of the energy consumption anomalies. However, the authors did not provide any information on the easiness of understanding of the information displayed using Sankey diagram.

In [6] the authors investigated the applicability of three different visualization techniques supplemented with anomaly scoring system to forming overall understanding on power usage and anomaly detection in power consumption. They assessed the Recursive Pattern, the Spiral view and the line charts. The Recursive pattern and the Spiral View are pixel–based visualization techniques that use coloring to encode numerical values in space filling layout, and differ from each other by the location of the time axis. The evaluation process of the visualization models consisted in the interviewing two experts from the company developing sensor networks measuring the power consumption for large buildings. They were shown the prototype and asked to give comments on it. In general, the experts liked the Recursive pattern and appreciated it as tool to form overview. The experts did not have possibility to apply it to solve tasks in order to assess the visual models on practice, no laboratory study was implemented.

In [8] authors proposed an approach to anomaly detection in HVAC system that is based on combination of three visualization techniques. They are RadViz visualization that is used to form daily operation patterns and detect suspicious deviations, matrix-based representation of the HVAC parameters that is constructed in the way that allows highlighting changes in values of parameters being analyzed and line charts. To validate the choice of these techniques, authors performed a set of experi-

ments that showed that these techniques form different visual patterns for normal and abnormal system functioning. However, the authors did not provide a thorough usability assessment of the approach that allows concluding that these particular models are easily comprehended by and useful for operators or security analysts.

Thus it is possible to conclude that visual analytics approaches presented above in the most cases lack usability assessment proving their efficiency in the HVAC or energy data analysis. There is a need to evaluate visualization techniques used in pattern detection and anomaly investigation process based on the feedback of the end users, i. e. managers of smart buildings and HVAC specialists.

In the conducted laboratory experiment we assessed the following visualization techniques: RadViz, matrix-based, and line charts as they are proposed in [8]. We decided to include line charts to the scenario of the laboratory study as a base line. This visualization technique is the most common way to represent data from sensors of any kind, and operators and analysts are used to it.

3   RadViz and matrix-based visualization of HVAC data

In this section we provide brief description of the RadViz and matrix-based visualization techniques used to analyze HVAC data and justification of their choice for anomaly detection.

The underlying idea that defined the choice of given visualization models consists in that the state of the HVAC system can be represented by a point in the multidimensional space. This assumption is justified by the fact that in general case, the functioning of the HVAC system is determined by the readings of the sensors monitoring the building's climate.

The goal of the RadViz visualization is to provide overview on how HVAC system functions, to highlight typical patterns and anomalies if possible. RadViz is a visualization algorithm that performs projection of the original multidimensional space into 2-dimensional space. The analyzed attributes are anchors or dimension nodes that are placed uniformly around the circumference of a circle. The object is represented by a point located inside the circle; the point is connected by n imaginary springs to the n respective anchors. The stiffness of each spring is proportional to the value of the corresponding attribute, and the point is located where the spring forces are in equilibrium.

Figure 1 shows the RadViz visualization of the HVAC data describing the state of the building during 8 working days. The selected color scheme is used to highlight the cyclic functioning of the system, and it is clearly seen that the points belonging to the time interval from 8.00 till 16.00 and partly to the time interval from 16.00 till 20.00 form rather dense region lying apart, allowing us to conclude that these points describe typical functioning of the HVAC system at working hours. Data points belonging to the intervals from 0.00 till 8.00 period form groups with linear shape that partly overlap with group of points belonging to the interval from 20.00 till 24.00, while the data points from time interval from 16.00 till 20.00 form scatter group of points also overlapping with points belonging to the interval from 20.00 till 24.00. This allow us

to conclude that there is a transitive period of the HVAC system functioning starting after 16.00 and finishing by 24.00; night mode of the HVAC system operation is described by a groups of points of linear shape.
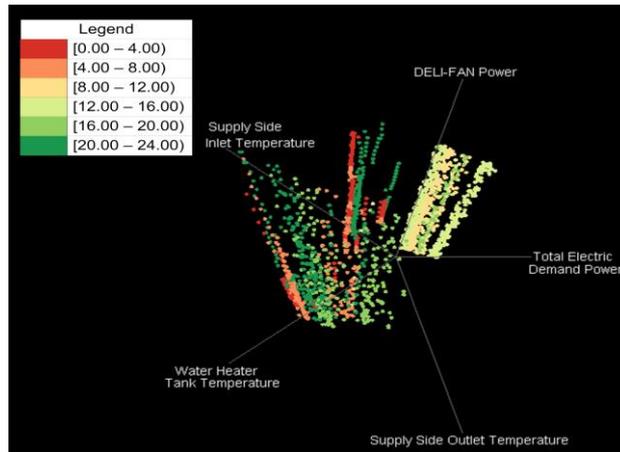

Fig. 1. RadViz presentation of the HVAC data for 8 working days

The RadViz does not provide any explicit information on attribute value, and the goal of the second matrix-based visualization model used in the approach is to provide generalized information on how the values of the HVAC parameters change and provide information on existing correlation between them.

The matrix-based presentation of the HVAC data displays averaged values for the attributes belonging to one data slice. In the approach the density-based clustering DBSCAN algorithm is used to construct data slices [9]. As it is initially designed to analyze static data without temporal attributes it was modified to consider the temporal order of data points [7].
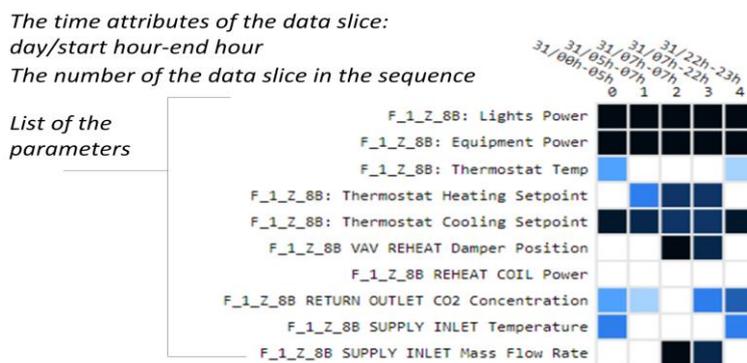

Fig. 2. The matrix-based presentation of the HVAC data

In the matrix-based presentation the rows correspond to the monitored system attributes, and columns represent data slices (Fig. 2). Each cell represents the average meaning of the attribute in the data slice. The color saturation is used to encode the values. The higher the value, the darker the matrix cell.

Figure 2 shows how the HVAC parameters changes during one day for the building corridor. It is clearly seen that the lights and equipment power are always on (parameters F_1_Z_8B Lights Power and F_1_Z_8B Equipment Power are almost dark.). The concentration of the CO_2 (parameter F_1_Z_8B RETURN OUTLET CO2 Concentration) decreases almost by 0 approx. at 7 o'clock and then increases. The decrease of CO_2 concentration is explained by the increase of the inlet air flow as the parameters F_1_Z_8B SUPPLY INLET Mass Flow Rate and F_1_Z_8B VAV REHEAT Damper Position increase significantly by the same time. But the subsequent increase of the CO_2 concentration can be explained by the starting of the work day.

4 Usability research of the HVAC data visualizations

4.1 Usability Experiment Scenario

There are many factors determining the choice and efficiency of the visualization technique applied to solve an analytical task. The type of visualization is determined both by data – amount of measured physical parameters, data sampling step, and by analytical task – comparison, filtering, structure or relation detection etc. Another important group of factors determining the efficiency of the visualization model is a purposefulness of the design and its elements, the information structural layout, the optimal use of colors, a rational approach to the data provided for accurate analysis and forecasting in the context of the situational perception of the operators.

To analyze the efficiency of the RadViz and matrix-based presentation for the HVAC data analysis we designed following laboratory study scenario.

First of all, we defined the goal of the laboratory study as follows: to define the efficiency of the selected techniques in data exploratory analysis aimed to detect anomalies. This enabled us to formulate following analytical tasks to be solved:

— finding and explanation of the repetitive patterns in HVAC data (search of groups and relations between parameters), and
— detection of anomalies in HVAC data and explanation of their origin (search of outliers).

When designing the usability study scenario, we kept in mind that participants of our survey may be limited in their free time, that is we tried to model the analysis task in such manner that it could be difficult enough but solvable in no more than 60 minutes. This decision also helped us not to overburden the participants and keep them focused.

To model the analysis task, we selected data set provided within VAST Challenge 2016 as test data to be explored. This data set is very close to a realistic one, it de-

scribes the functioning of 3 storied building with dozens of zones controlled by HVAC systems. Each zone consists of different rooms having different purposes – offices, conference halls, meeting rooms, cafes and server room. Thus, HVAC system functions differently in each controlled zone. The data are sampled every 5 minutes during 14 days. There is a set of anomalies in HVAC system that are in major cases connected with wrong setting points for temperature parameters. However to be able to finish the analysis task in a hour we limited the data to data describing one zone only; and selected 5 days: 2 days with normal system's behavior, 2 days of easily detectable abnormal behavior and 1 day of hardly detectable abnormal behavior;

The latter allowed us to have subtasks of different complexity level. We also included few parameters with non-informative data.

Thus, we selected data for 1 controlled zone (Floor 3, Zone 9), 5 working days (31.08, 01.09, 06.09, 08.09, and 09.09) and 9 registered parameters (Lights Power, Equipment Power, Thermostat Temp, Thermostat Heating Setpoint, Thermostat Cooling Setpoint, VAV Damper Position, Return Outlet CO_2 Concentration, and Supply Inlet Temperature, Supply Inlet Mass Flow Rate).

For every of the 5 days we produced 3 different visualizations: RadViz, matrix-based visualization, and linear charts. We decided to include them to the scenario of the laboratory study as a base line. This visualization technique is the most common way to represent data from sensors of any kind, and operators and analysts are used to it.

Apart of charts with different visualization models we also provided a scheme of a hypothetical room with the registered parameters, it was used to demonstrate the main dependencies between parameters.

To measure the comprehension level of data presented by visual models we designed a questionnaire that contained questions of 3 categories of difficulty:

— easy questions: 4 questions relating general description of HVAC system working mode;
— medium complexity questions: 3 questions concerning a qualitative assessment in changes of system parameters (for normal and abnormal days), and
— difficult questions: 2 questions concerning hypothesis statement, the subjects were offered to formulate hypotheses about cause-and-effect relationships and regularities in the data.

We selected following quantitative metrics to assess the efficiency of the visualization techniques: 1) frequency of use, 2) time spent on completing tasks (on each task and in general). To understand the ways respondents use different visualization models we asked respondents to indicate the role of each visualization model in data analysis process (primary or secondary). We also were interested in receiving feedback on subjective preferences of the participants, and proposed them to evaluate following metrics: 1) visual models clarity, 2) visual models utility and 3) general interest in solving tasks. These parameters were assessed on five point scale, where 1 means worse, and 5 – excellent.

4.2   Experiment Implementation

The experiment was conducted in 2 stages: the trial and main stages.

The trial phase was used to check the experiment design. Since this stage was designed as qualitative, we decided that a minimum but carefully selected number of respondents will show the main advantages and disadvantages of each visual model that could be further used to improve experiment handouts and would help us better understand the aspects of user interaction with data visualization. Two postgraduate students (23 years old in average) with no special training and experience in HVAC systems participated in the trial stage. They worked together solving the given tasks. The subjects worked in the usability testing room, the experimenters were observing the participants from the next room behind the Gesell's mirror, the participants and the experimenters were communicating via a speakerphone. The subjects worked with the given information, visualizations, and question forms on a computer. Based on the results of the trial phase, the visualizations, information provided to the subjects, as well as questions were slightly modified.

Two groups of respondents participated in the main stage of the experiment, the total number of participants was 8 persons. Four subjects had special training and experience in cyber-physical system operation: three persons had experience in working with refrigeration equipment, while another one – with magnetic resonance tomography. Among them there were 3 males and 1 female, with average age 35 years old. Other four subjects had no special training and no experience as system operators, they were students of ITMO University master program "Multimedia Technologies, Design and Usability" (they were all females whose average age was 23 years old). We wanted to compare two groups of subjects with different level of experience in operating cyber-physical systems and application of data visualizations to solve analytical tasks in order to identify differences in their understanding of the three types of visualizations, and to determine how the experience of cyber-physical system operators affects their success in tasks solution.

During the main stage the participants worked individually. Instead of the work on a computer, they were given printed handouts with general information, visualizations with descriptions, the questionnaire form.

At the beginning of the experiment the subjects were asked to read the general information. The general information was based on VAST Challenge description and included:

— brief description of the HVAC system installed;
— list of the 9 parameters with their description;
— a schema of the controlled zone to be analyzed.

After reading the general information, the subjects were introduced to the investigated data visualizations and their description. The subjects were given 15 minutes to understand them and afterwards they could ask the questions.

Then the subjects were given a questionnaire form to fill in and instructed that they may use any information they have in any order. We asked respondents to spend 30

minutes maximum on tasks. When they completed the questionnaire, we thanked them and asked for a feedback using a subjective preference form.

4.3 Experiment Results and Discussion

The majority of the participants answered all the questions given: they made their assumptions about the normal/abnormal days, abnormal data, and explained possible causes of anomalies. All operators put forward versions of all the questions asked, and two non-operators left empty fields for answers to one question of medium complexity and one complicated question.

A group of non-operators solved the tasks faster than the other group, spending on average 32 minutes to solve the test. It took the operators more time, averaging 53 minutes. Perhaps, this is explained by the fact that operators approached more responsibly to the survey, and general interest in solving task was higher in their group. The average score for the interest in the operators group is equal to 4.25 points in 5 item scale (two of them ranked interest level as average, and two of them – as high). The average rate of interest of non-operators is 3 (two of them showed average interest, two showed low interest), indicating lower level of motivation and effort in this group of respondents.

Interestingly, that the most used visual model among all respondents regardless their experience was the matrix-based visualization. RadViz visualization was the least used one (Fig. 3).

It is clearly seen that the matrix-based presentation is the most frequent visualization technique in both groups of subjects - operators used it more often than non-operators. Linear charts were more frequently used by test subjects with experience in operating cyber-physical systems. This fact could be explained by that participants-operators of refrigeration equipment interact exactly with this type of visualization in their professional activity, and, therefore, it is more usual for them. RadViz visualization was practically not used by non-operators and was not used by operators at all, probably because of difficulties in interpretation. Experienced operators did not try to discover any useful information in it considering limited time for solving tasks and not wasting time on it.
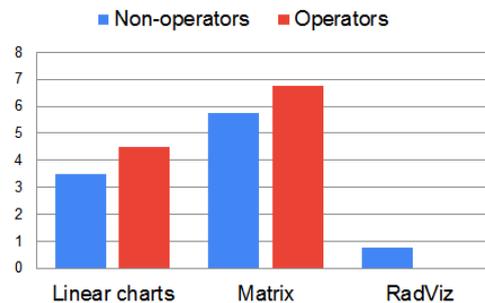

Fig. 3. Frequency of visualization use.

These results are consistent with subjective assessments of the laboratory study participants. In general case linear chars and matrix-based visualization received almost similar scores in clarity and utility (Fig. 4). However, the clarity of linear charts was ranked slightly higher than matrix-based visualization by operators. Operators noticed that the linear charts are more understandable to them, as they are used to working with such data visualization.

RadViz technique received the lowest scores in utility and clarity. Surprisingly, operators ranked it higher than non-operator respondents, considering the fact they did not use it in task solution. Probably, the reason lies in better understanding of the linear charts and matrixes due to professional expertise and limited time given for task solution, while it was not clear to them when to apply RadViz visualization in the task suggested. Nevertheless, one of operators pointed to its important role in situations of rapid assessment of information and decision-making.
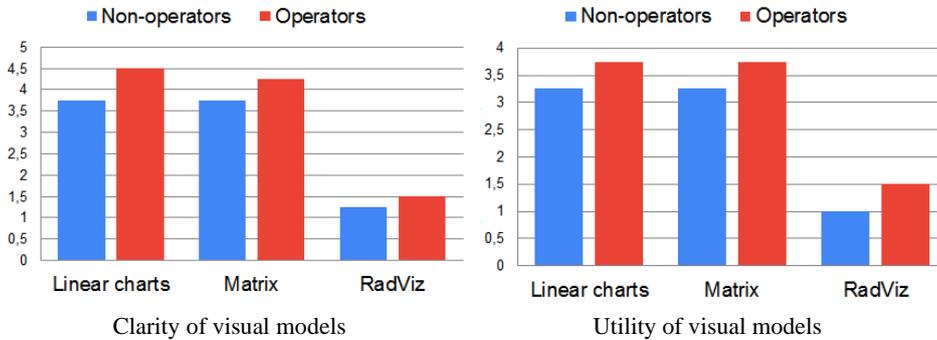

Clarity of visual models   Utility of visual models

Fig. 4. Subjective assessments of visualization clarity and utility on a five-point scale.

At the beginning of the study, the hypothesis was advanced that RadViz visualization is the most informative in rapid analysis to get a general idea of data behavior and possible deviations. In the course of the experiment, it was practically not used by the subjects in the process of finding answers to tasks. The results of the experiment suggest that the subjects of both groups encountered difficulties in interpreting this visualization due to insufficient elaboration. Subjects claimed that they were confused by the large number of axes and insufficient size of the information presented. Only one subject with experience in operating cyber-physical systems indicated the potential benefits of using RadViz in data analysis over a short period of time. It can be concluded that RadViz requires a more accurate elaboration in its visual representation (scaling font and recorded points), as well as longer and more intense training of respondents to comprehend this type of data visualization.

The presentation of data in the form of linear charts is more familiar, understandable and useful for experienced operators of cyber-physical systems. The advantage of this visualization is its high level of detail, which is useful for detailed interpretation about the system's behavior. Linear chart visualization was the most reworked one after the trial experiment, despite this it appeared to be unclear for some subjects because of its small size represented on A4-size paper (mentioned by two subjects). The main disadvantage of the graph visualization was the lack of the vertical lines depict-

ing the time intervals, which impaired the temporal data interpretation (the fact was mentioned also by two subjects).

Matrix-based data presentation was most often used to answer questions from subjects in both groups, and was also highly rated for its clarity and usefulness. Studying it takes more time than studying RadViz and it is not as detailed as linear charts are, but based on the experiment results, we may assume that if we restricted our subjects to use only one visualization in solving the modeled task, the matrix-based visualization would show the highest efficiency. Subjects did not comment on the presentation of this visualization, confirming the fact that it is uniform and informative enough to answer questions in this experiment.

Thus, the analysis of the experimental results allowed us to rank these visualization techniques as follows.

— Linear charts provide a very detailed view of the system's behavior, but require too many resources, they also causes some technical difficulties in their representation. We refer to such techniques as "slow but detailed" techniques.
— Matrix-based visualization can be considered as a technique that can illustrate both general patterns of system behavior and details of data changes. It is possible to think of them as "in-between" methods.
— RadViz is perhaps a promising technique for providing fast insight in data, however, it requires a solid research in enhancing its visual representation and comprehension by users. This type of visualization techniques may constitute "fast but superficial" methods to analyze data.

Thoroughly elaborated design of these models as well as interaction techniques linking them may result in producing effective analytical dashboard for HVAC operators and security analysts.

5   Conclusions

This experiment gives an idea of advantages and disadvantages of three types of visualizations – Linear Charts, Matrix-based visualization and RadViz – in the context of solving HVAC data exploratory analysis task.

Visualization-driven approaches may significantly increase the efficiency of the HVAC data if the visualization models are appropriate to the task solved and easily comprehended by the end users. The usability testing is a good way to evaluate the visualization technique efficiency. The paper presents the results of the laboratory study of the three visualization models designed to support HVAC data analysis – RadViz, matrix-based visualization and linear charts. The experiments showed, that one should use the "fast but superficial" methods (for example, RadViz) in order to quickly obtain the screening results (familiarization with the data structure and behavior and possible abnormalities in the data). To get some detailed information about the data behavior and cause-effect dependencies, one should use "slow but detailed" methods such as linear charts. For defined tasks "in-between" methods can be more efficient, as they are slower than the "fast but superficial" methods in giving screen-

ing results and they cannot give as detailed information as "slow but detailed" ones do. For complex tasks involving iterations of obtaining the quick screening results and getting the detailed information on the highlighted data, one should combine both "fast but superficial" and "slow but detailed" techniques, complementing it with "in-between" visualization techniques. Thus, it is possible to conclude that proposed approach can be used in the HVAC data analysis but still further evaluation, e.g. comparison with conventional visualization models, is required.

Further researches on the efficiency of data visualizations obtained from HVAC-type cyber-physical systems may include more thorough analysis of the test subjects' answers to the degree of their correctness and elaboration of hypotheses. There is also a need to increase the number of subjects involved in cyber-physical system operation for quantitative research that includes statistical analysis of data.

Acknowledgements

References

1. Ciholas P., Lennie A., Sadigova P., Such J. M.: The security of smart buildings: a systematic literature review (1), (2019).
2. Murugesan L., Hoda R., Salcic Z.: Design criteria for visualization of energy consumption: A systematic literature review. Sustainable Cities and Society (18), pp. 1–12, (2015).
3. Truong H., Francisco A., Khosrowpour A., Taylor J. E., Mohammadi N.: Method for visualizing energy use in building information models. Energy Procedia (142), pp. 2541–2546, (2017).
4. Palm J., Ellegård K.: Visualizing energy consumption activities as a tool for developing effective policy. International Journal of Consumer Studies (35), pp. 171–179, (2011).
5. Abdelalim A., O'Brien W., Shi Z.: Development of Sankey diagrams to visualize real HVAC performance. Energy and Buildings (149), pp. 282–297, (2017).
6. Janetzko H., Stoffel F., Mittelstädt S., Keim D. A.: Anomaly detection for visual analytics of power consumption data. Computers & Graphics (38), 27–37, (2014).
7. Novikova E. Bestuzhev M. A.: The visualization-driven approach to the analysis of the HVAC data. In: Intelligent Distributed Computing XIII. IDC 2019. Studies in Computational Intelligence, vol. 868, pp. 547-552, (2020).
8. Novikova E., Bestuzhev M., Kotenko I.: Anomaly Detection in the HVAC System Operation by a RadViz Based Visualization-Driven Approach. In: Katsikas S. et al. (eds) Computer Security. ESORICS 2019 International Workshops, CyberICPS, SECPRE, SPOSE, and ADIoT CyberICPS, Lecture Notes in Computer Science, vol 11980. Springer, Cham (2019).
9. Ester M., Kriegel H.-P., Sander J., Xu X.: A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: Jiawei Han, Usama M. Fayyad (eds) Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96), vol. 3, pp. 226–231, (1996).