Систематизація даних щодо енергоспоживання будівель різного призначення дозволяє дослідити процеси з позицій ефективного використання енергоресурсів для забезпечення комфортних умов. Це обумовлює необхідність удосконалення існуючих або пошуку нових підходів до аналізу даних щодо енергоспоживання різних будівель.

На основі дослідження процесу підготовки даних про енергоспоживання будівель запропоновано методику первинного аналізу. Враховуються призначення будівлі та способи отримання даних, інформація щодо показників абсолютного й питомого електро- та теплоспоживання, показники температур внутрішнього і зовнішнього повітря. На прикладі енергоспоживання будівель навчального закладу виконана апробація розробленої методики початкового аналізу даних.

Дослідження енергоспоживання будівель навчального закладу дозволили встановити кількісну зміну енерговитрат та температур зовнішнього й внутрішнього повітря при переході від річних даних до даних опалювального періоду. Аналіз діаграм розмаху наглядно демонструють діапазони зміни показників енергоспоживання, їх середні значення та викиди, що вказують на збільшену понад норму витрату енергоресурсів досліджуваними об'єктами в перехідні періоди.

Для наочної оцінки ефективності використання одиниці теплової енергії одиницею площі різних квартир з різними способами обліку та регулювання теплоспоживання розроблено web-базоване програмне рішення, у якому порівнюються показники теплоспоживання окремих квартир з максимальною потребою на опалення відповідно до нормативних документів.

Отримані результати слугують основою для створення прикладних інформаційних рішень для муніципального енергоменеджменту

Ключові слова: теплове забезпечення будівель, аналіз даних енергоспоживання, підготовка даних, моніторинг енергоспоживання будівель

PREPARATION AND PRELIMINARY ANALYSIS OF DATA ON ENERGY CONSUMPTION BY MUNICIPAL BUILDINGS

A. Perekrest
PhD, Associate Professor
Department of automation
and computer-integrated technologies*
E-mail: pksg13@gmail.com

O. Chornyi
Doctor of Technical Sciences,
Professor, Director of Institute
Institute of Electromechanics, Energy Saving
and Automatic Control Systems*
E-mail: alekseii.chornyi@gmail.com

O. Mur
Department of automation
and computer-integrated technologies*
E-mail: alexanndra.moore@gmail.com

A. Nikolenko
PhD, Associate Professor, Head of Department**

V. Kuznetsov
PhD, Associate Professor**
E-mail: wit_jane2000@i.ua

Y. Kuznetsova
Senior Lecturer
Department of humanitarian,
fundamental and general engineering disciplines
Institute of Integrated Education***
E-mail: wit1975@i.ua
*Kremenchuk Mykhailo Ostrohradskyi
National University
Pershotravneva str., 20, Kremenchuk, Ukraine, 39600
**Department of the electrical engineering
and electromechanic***
***National Metallurgical Academy of Ukraine
Gagarina ave., 4, Dnipro, Ukraine, 49600

1. Introduction

At present, much attention is paid to the problems of efficient use of energy resources and energy efficiency. Measures that are implemented with the aim of energy-saving are not typically effective enough. This is related to the absence of complete information about objects that utilize energy and about the specificity of their operation, the lack of the required amount of funds to conduct activities aimed at increasing energy efficiency of buildings, etc.

Solving a task on energy efficiency implies accounting of energy resources, analysis of energy-consuming modes, as well as decision-making and management. Practical implementation is, as a rule, based on constant monitoring and parameterization of the system, which leads to the accumulation of a large amount of historical data on energy consumption, obtained from the nodes of commercial accounting and electronic weather-dependent regulators.

Solving the problems on energy efficiency and energy saving requires a comprehensive approach, which predetermines the need to improve already existing approaches,

or to search for the new ones, in order to analyze data on energy consumption [1–3]. Processing of large volumes of data predetermines the application of specialized algorithms and methods of statistical and intelligent analyses [4–6].

Data collection, preparation and conducting the initial analysis, make it possible to obtain valuable information on energy consumption in order to build a model using methods of machine learning. An initial analysis exerts a significant impact on the adequacy of the models being constructed. There are objective difficulties associated with the specificity of a subject area, the combination of data from multiple sources and the lack of a unified procedure for data analysis, whose application results could be implemented into existing systems.

The development of information technology has led to an increase in the number of sectors where methods of machine learning are employed. Among them is the management of power consumption and the prediction of an energy resource cost. Methods of machine learning help in the analysis of parameters for a power management system and support making effective decisions by energy managers.

There is a need to improve already existing approaches, or to search for the new ones, in order to analyze data on energy consumption, with the aim of making decisions directed at increasing the efficiency of energy resources use.

Implementation of decisions related to the automated monitoring and control over heating systems makes it possible to reduce the overall consumption of thermal energy and to accumulate large amounts of information on the functioning of the system and on the decisions made by an energy manager. That solves the tasks associated with the need to ensure regulatory sanitary and hygienic conditions in the premises that are heated, taking into consideration the influence of external factors.

2. Literature review and problem statement

Paper [7] considered the possibility to predict heat consumption in the system of centralized heat supply based on the Elman neural network using a genetic algorithm, which makes it possible to optimize the weights of models. Research [8] confirms that the results obtained by using neural networks are more precise compared with the ensemble random forest method when solving a problem on predicting energy consumption. However, both models can be used for an hourly prediction. The authors of [9] analyzed characteristics of energy consumption and factors affecting it, based on data on the operation of shops' facilities. Since the authors employed a variety of methods for intelligent analysis of a significant volume of data, the findings obtained make it possible to optimize construction costs and business equipment costs based on the built prediction model.

Despite the practical significance of such results, the authors failed to sufficiently consider the process of implementation of the obtained results into existing decision support systems. The description of the actual process of analysis has not been sufficiently documented. In addition, having been developed only to examine data on the energy consumption by shops, it implies the use of specific factors in its assessment. Studies into the accuracy of models that employ artificial neural networks for a short-term forecasting of supplying heat to objects within a social-budget-funded sector have found that the best results had been demonstrated by a model of the NARX type [10]. Because the data under con-

sideration are typically of a mixed type, incorrectly structured, and/or have missing values, paper [11] synthesized the models for data imputation based on several methods of machine learning. The authors compared them with a standard method that is included in most statistical packages.

Paper [12] presented a method for diagnosing energy efficiency based on a limited amount of data acquired from the accounts for electricity and data about weather conditions. The obtained solutions provide an opportunity, based on the results of a multi-parametric regression, to determine volume of the resource consumed by each system, as well as detailed information about the physical properties. The authors also proposed ways to obtain the necessary data. However, despite the advantages of the method, the issue of data preparation and preliminary analysis remains unexplored.

The development of aspects related to research and, in fact, to datasets about energy consumption, is outlined in paper [13]. The authors categorized existing datasets according to the three defined strategies for the generation and development of the catalogue of available data on energy consumption. The authors of this work confine themselves to collecting data and their grouping, as well as to the general presentation of their classification apparatus.

Researchers do not pay enough attention to the issue of data collection, preparation and conducting an initial analysis, which would make it possible to acquire information about power consumption prior to constructing a model and could significantly affect the adequacy of the model. Therefore, there is a need to build a unified methodology to examine data on the energy consumption in buildings with different functional purpose.

3. The aim and objectives of the study

The study that we conducted set the goal of improving the quality of analysis of energy consumption by municipal buildings through automating the procedure of processing primary data that were acquired from commercial accounting nodes or from the installed systems for monitoring power consumption.

To achieve the set aim, the following tasks have been solved:

– to identify the features of approaches to examining data on energy consumption, to adapt existing methodologies in accordance with the characteristics of the subject area and to construct a procedure for initial analysis of data on energy consumption by buildings;

– to substantiate the structure of information flows of data on the energy consumption by different buildings, to implement a set of software solutions for retrieving data from different sources while providing the necessary level of their informativeness and quality, and to test the procedure for initial analysis of data on energy consumption by buildings of the educational institution;

– to develop a web-based, application software in order to visualize an analysis of thermal energy consumption by different apartments with different techniques for accounting and regulation of heat consumption.

4. Materials and methods to examine data

4. 1. Information data flows

A sector that consumes a significant portion of energy resources is the system of heat supply to facilities and resi-

dential objects in cities: residential fund, buildings of social, educational purpose, etc.

The objects whose data were investigated were conditionally divided into four groups: the first group is composed of buildings with a heat meter, the second group includes buildings without a heat meter, the third group consists of private houses, the fourth group includes administrative premises.

We selected as administrative premises the academic buildings of Kremenchuk Mykhailo Ostrohradskyi National University (KrNU, Ukraine). KrNU has developed and installed an automated system for operative control and management (ASMU) over thermal points at separate academic buildings. The system is composed of standard industrial equipment. The developed software has made it possible to execute current control over temperature modes of heating systems and to operatively change setpoints for a heat carrier's temperature depending on the outside temperature and working conditions at buildings [14]. The result of ASMU work to handle the heat consumption by academic premises is the availability of data about actual operational modes of heating systems, as well as on energy consumption by separate buildings.

An objective and accessible source for obtaining initial data on buildings without an installed ASMU is the bills for utilities.

As regards the administrative and budget-funded buildings, the consumption of thermal energy is based on a building's meter, of electricity – on a building's meter, the outer volume of buildings and the cost of thermal energy are known.

An analysis of the bills has revealed that for buildings with a heat meter there are available data on the area of an apartment, the cost of thermal energy, share of an apartment in the heating area of a building, consumption based on an apartment's meter, the building's meter, and the amount to be paid.

Similarly, for apartments in buildings without heat meters there are available data on the area of an apartment, the tariff, and the charged amount to pay for a service.

An analysis of the bills for private houses has showed that known data include the volume of consumed gas, heating area, the cost of gas.

4. 2. Procedure for initial analysis of data

Data Mining is a process whose goal is to find significant correlations, models, and trends, by processing a large volume of data using a variety of statistical and mathematical methods. It is typical to employ the sectoral standard for data examination, CRISP-DM. In accordance with it, a life cycle must be composed of six phases: understanding business processes, understanding the data, data preparation, modelling, evaluation, implementation.

When analyzing data on energy consumption, there is a need to solve certain business tasks:
– Which buildings are the most energy efficient?
– Which premises are the most comfortable for people?
– Which buildings are the most investment attractive?
– Which buildings need priority modernization?
– What is the quality of the provided heating service?

Therefore, it is advisable, when examining data, to be guided by one of the existing analysis methodologies. Given that analytical data processing on energy consumption is a typical project within Data Mining, it was decided, based on CRISP-DM, taking into consideration the specificity of the subject area, to construct a unified procedure for initial analysis of data on energy consumption.

The constructed procedure implies the following stages:
– collecting source data: request to access data from a list of the project's resources and downloading them;
– description of data: examination of the basic, or the "surface", properties of the collected data and description of results: the amount of data, types of values, encoding schemes;
– data examination: analysis of "grey" data by using queries, visualization, and reporting;
– checking the quality of data: existence of encoding errors and missing values;
– data selection: selection of elements and attributes;
– conducting a series of syntactic transformations due to the specificity of functioning of the selected tools for analysis;
– data selection and merge: merging data tables, which contain different or the same information about the same objects;
– generation of attributes: overall power consumption, specific heat-, electricity-, and overall energy consumption, generated attributes reduced to the regulatory values for external and internal air temperature, the number of degree-days, a mark of the heating season (if necessary);
– analysis of general descriptive statistics for a dataset examined at a given stage;
– check for normality of distribution: conducting a Shapiro-Wilk test; if required, undertaking a separate study into asymmetry and excess taking into consideration the sample size;
– conducting a correlative analysis for the quantitative estimation of relationship among data;
– building boxplots or a nuclear density estimation for comparing the probability distribution between the groups of data and an analysis.

In accordance with the constructed procedure, we provide the following descriptive statistics: arithmetic mean, mean square deviation, median, minimum, maximum, asymmetry, excess, standard error of the mean.

To check the normality of distribution, we perform the Shapiro-Wilk test. A null hypothesis that is tested using the Shapiro-Wilk test, is stated as follows "The examined sampling is derived from a general totality that has a normal distribution". If the error probability, obtained using the test, turns out to be less that the predefined level of significance (for example, 0.05), the null hypothesis is rejected.

Considering the results on testing the normality of distribution, we define a correlation method: a Pearson coefficient is used for normally distributed data, the non-parametric Spearman coefficient is applied for data whose distribution is different from normal.

5. Results of research into energy consumption of buildings

5. 1. Structures of information data flows on energy consumption by different buildings

In order to simplify the data and make them more convenient for preparation, we developed structures of information flows for different objects.

Thus, for a building with a heat meter (Fig. 1), we first determine the total heating area:

$$A_t = \frac{A_{ap} \cdot 100\%}{dA_{ap}}, \qquad (1)$$

where A_t is the total heating area, m²; A_{ap} is the area of an apartment, m²; dA_{ap} is the share of an apartment in the overall heating area.

We then determine specific heat consumption:

$$q = \frac{\Delta E_b \cdot 1163}{A_b}, \qquad (2)$$

where q is the specific heat consumption of a building, kW·hour/m²; A_b is the building's area, m²; ΔE_b is the consumption based on a meter, Gcal.

Energy efficiency class is:

$$CEE = \frac{q - E_{\max}}{E_{\max}}, \qquad (3)$$

where CEE is a coefficient of energy efficiency; $E_{\max}$ is the maximum specific need for heating, kW·hour/m²; q is the specific heat consumption, kW·hour/m².

Heating tariff is:

$$T = \frac{\Delta E_b \cdot C_{te}}{A_t}, \qquad (4)$$

where T is the heating tariff, UAH/m²; C_{te} is the cost of thermal energy, UAH/Gcal; A_t is the total heating area, m²; ΔE_b is the consumption based on a meter, Gcal.
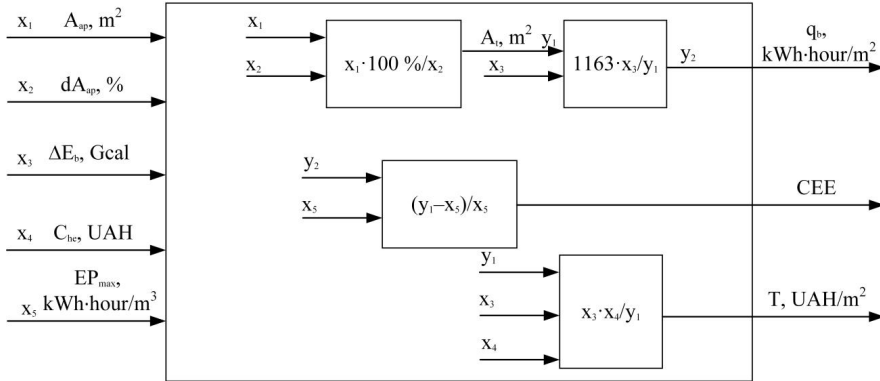
Fig. 1. Structure of information flows for buildings with a heat meter

For buildings without a heat meter (Fig. 2), we determine the amount of heat energy to heat premises:

$$\Delta E_{ap} = \frac{Am_{ap}}{C_{te}}, \qquad (5)$$

where ΔE_{ap} is the amount of heat energy used to heat a premise, Gcal; Am_{ap} is the amount to be paid, UAH; C_{te} is the cost of heat energy, Gcal.

The following stages imply determining specific heat consumption (2) and the coefficient of energy efficiency (3).

Since the overall heating area is unknown, the indicators are determined for an apartment (Fig. 3).

In order to determine indicators of heat consumption by private houses with a natural gas meter (Fig. 3), we first must reduce initial data to stan-

dard conditions by multiplying the volume of gas consumed by its respective coefficient:

$$V_{st\_c} = V_{w\_c} \cdot C_{st}, \qquad (6)$$

where V_{st_c} is the volume of gas reduced to standard conditions, m³; V_{w_c} is the volume of gas under working conditions, m³; C_{st} is the coefficient to reduce gas to standard conditions.

The application of one or another coefficient depends on the location of a gas meter, region or oblast, and a corresponding month. Because a natural gas counter registers the total volume of gas used by consumers, there is a task to determine the volume of gas used to heat a building:

$$V_h = V_{st\_c} \cdot V_{h\_n}, \qquad (7)$$

where V_{h_n} is the volume of gas that is used for household needs. V_h is the volume of gas used for heating, m³; V_c is the volume of gas reduced to standard conditions, m³.

We then determine the absolute heat consumption to heat a private house:

$$\Delta E_b = \frac{GCH \cdot V_h}{10^6}, \qquad (8)$$

where ΔE_b is the absolute heat consumption to heat a house, Gcal; GCH is the gas combustion heat (caloric content), kcal/m³; V_h is the volume of gas used for heating, m³.

Caloric content is the primary indicator of quality of natural gas. The higher the caloric content, the less the amount of gas that is required to meet specific needs. The values for caloric content are fixed by service providers and are usually available for use.

Heating tariff is:

$$T = \frac{V_h \cdot C_g}{A_b}, \qquad (9)$$

where T is the heating tariff, UAH/ m²; C_g is the cost of gas, UAH/m³; A_b is the building's area, m²; V_h is the volume of gas used on heating, m³.

The obtained results make it possible to determine specific heat consumption and an energy efficiency class (2)–(3).
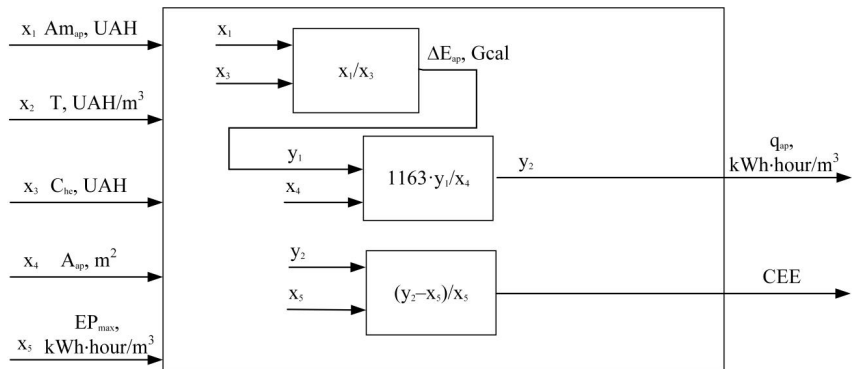

Fig. 2. Structure of information flows for buildings without a heat meter

For the administrative and budget-funded facilities (Fig. 4) we perform calculation similar to buildings with a heat counter (2–4), and thus derive values for specific heat consumption, tariff, and the class of energy efficiency.
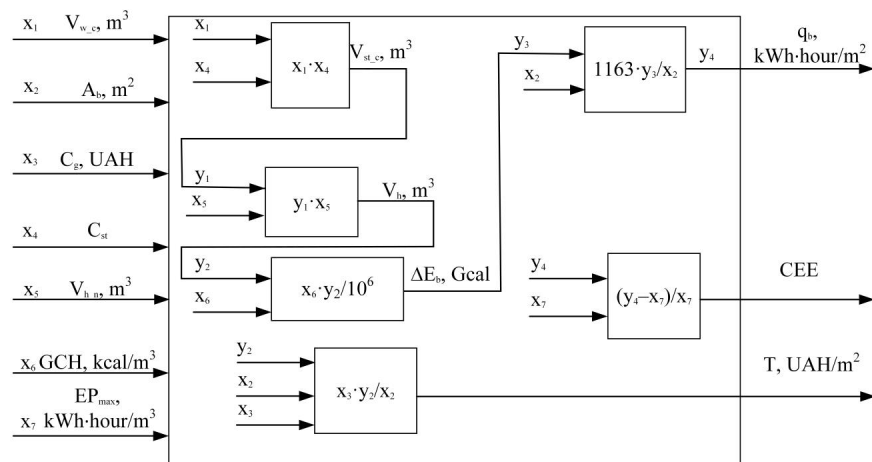

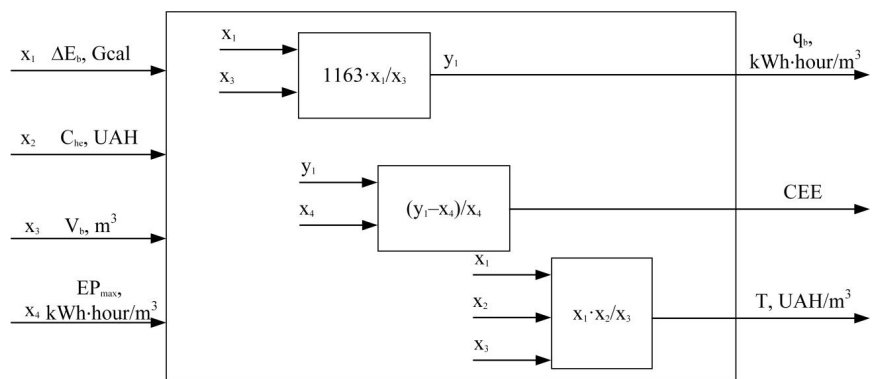Fig. 3. Structure of information flows for private houses

Fig. 4. Structure of information flows for administrative and budget-funded buildings

5. 2. Verification of the procedure for initial analysis of data

In order to verify the devised procedure, we employed data on the energy consumption by seven academic buildings at KrNU over the period from 2012 to 2016.

For the present study, we collected three datasets:

– Energy consumption by buildings of the educational institution over the period from 2012 to 2016, indicating the number of a building, month of the year, consumption of heat energy (expressed in Gcal), electricity consumption (expressed in kW·hour) (Table 1).

The volume of data: columns – 5, rows – 420.

The format of data: the number of a building is the categorical variable; the rest of the data are numeric.

Table 1

The encoding scheme of the set "Energy consumption and indoor temperature"

Number of a building	k
Month	month
Year	year
Heat consumption, Gcal	E_Gkal
Electricity consumption, kW·hour	W_kWh
Average monthly indoor temperature, °C	Tin_C

– Volume of the heat load, indicating the number of a building, the volume of heat load (expressed in Gcal/hour) and a building's volume (Table 2).

The volume of data: columns – 3, rows – 7.

The format of data: the number of a building is the categorical variable; the rest of the data are numeric.

– The average monthly ambient temperature over the period from 2012 to 2016, indicating the year, month, and the average monthly temperature (Table 3).

The volume of data: columns – 3, rows – 60.

The format of data: all data are numeric.

Table 2

The encoding scheme of the set "Volume of heat load"

Number of a building	k
Volume, m³	V_m3
Volume of heat load, Gcal/year	Qop

Table 3

The encoding scheme of the set "Temperature"

Year	year
Month	month
Average monthly ambient temperature, °C	Tout_C

The initial study is limited to data over 2012–2016 (420 elements), which is why we must adjust the filters so that data from other periods are not in the selected data.

The examined datasets do not contain mistakes and missing values, and the encoding schemes converge. After checking the quality, we shall proceed to the data preparation stage.

All of the above attributes were included in the dataset. Since the source collected data on energy consumption are represented in the first normal form, we made a transition to the third normal form, in order to eliminate transitive dependences. The result of merging the three tables is the unified table with monthly data on each building over 2012–2016, which contains the attributes selected at the previous steps of analysis.

In order to evaluate the initial data, there is a need in additional parameters for a comparative analysis of different objects. To perform a comparative analysis of the energy consumption by different buildings it is advisable to use the same indicators, for example, specific heat consumption ($q1$, kW·hour/m³), specific electricity consumption ($q2$, kW·hour/m³), etc.

The normative value for degree-days during heating period (DDHP):

$$Dd_N = (T_{in} - T_{a\_n\_am}) \cdot Z_N, \tag{10}$$

where Dd_N is the normed number of degree-days; T_{in}=20 °C is the indoor temperature; $T_{a\_n\_am}$=−0.8 °C is the average

normed ambient temperature; Z_N=180 is the normed duration of a heating period.

The actual value for degree-days of a heating period (DDHP):

$$Dd = (Tin\_C - Tout\_C) \cdot Z_{act}, \qquad (11)$$

where Dd is the actual number of degree-days; Tin_C, °C is the indoor temperature; Tout_C, °C is the average actual outdoor temperature; Z_{act} is the actual duration of a heating period.

A DDHP coefficient is:

$$K_{Dd} = \frac{Dd_N}{Dd}, \qquad (12)$$

where K_{Dd} is a DDHP coefficient; Dd is the actual number of degree-days; Dd_N is the normed number of degree-days.

Absolute heat consumption in kW·hour:

$$E\_kWh = E\_Gkal \cdot 1163, \qquad (13)$$

where E_kWh is the absolute heat consumption, kW·hour; E_Gkal is the heat energy consumption, Gcal.

Specific heat consumption is:

$$q1 = \frac{E\_kWh}{V}, \qquad (14)$$

where $q1$ is the specific heat consumption, kW·hour/m³; E_kWh is the heat energy consumption, kW·hour; V is the volume of a building, m³.

Specific heat consumption is:

$$q2 = \frac{W\_kWh}{V}, \qquad (15)$$

where $q2$ is the specific electricity consumption, kW·hour/m³; W_kWh is the consumption of electrical energy, kW·hour; V is the volume of a building, m³.

Total energy consumption is:

$$Esum\_kWh = E\_kWh + W\_kWh, \qquad (16)$$

where Esum_kWh is the total energy consumption, kW·hour; W_kWh is the consumption of electrical energy, kW·hour; E_kWh is the consumption of heat energy, kW·hour.

Specific total energy consumption is:

$$q3 = \frac{Esum\_kWh}{V}, \qquad (17)$$

where $q3$ is the specific total energy consumption, kW·hour/m³; Esum_kWh is the specific energy consumption, kW·hour; V is the volume of a building, m³.

Specific heat consumption, reduced to the normative values for outdoor and indoor air temperature, is:

$$q1t = q1 \cdot K_{Dd}, \qquad (18)$$

where $q1t$ is the specific heat consumption taking into consideration temperatures, kW·hour/m³; $q1$ is the specific heat consumption, kW·hour/m³; K_{Dd} is a DDHP coefficient.

Specific electricity consumption, reduced to the normative values for outdoor and indoor air temperature, is:

$$q2t = q2 \cdot K_{Dd}, \qquad (19)$$

where $q2t$ is the specific electricity consumption taking into consideration temperatures, kW·hour/m³; $q2$ is the specific electricity consumption, kW·hour/m³; K_{Dd} is a DDHP coefficient.

Specific total energy consumption, reduced to the normative values for outdoor and indoor air temperature, is:

$$q3t = q3 \cdot K_{Dd}, \qquad (20)$$

where $q3t$ is the total specific energy consumption taking into consideration temperatures, kW·hour/m³; $q3$ is the specific total energy consumption, kW·hour/m³; K_{Dd} is a DDHP coefficient.

In the case of analysis of energy consumption by buildings, in the transition from general data to data on the heating and non-heating periods, we convert the sample on consumption data, where each entry corresponds to monthly indicators, into a new sample, where each building is matched with a series of sorted records with characteristics for a separately taken heating (non-heating) period. In order to analyze seasonal energy consumption, we form a unified new attribute – a tag for a heating season.

Next, we convert the table with data on monthly energy consumption by buildings of the educational establishment, where each entry corresponds to one month, into a new table. In it, each consumer of energy is matched with a single record with general characteristics of his/its energy consumption (for example, the total amount of electricity consumed over a heating period, the number of degree-days over this period, etc.).

Fig. 6 shows charts for general data correlation matrices (Fig. 5), monthly data on heating (Fig. 6) and non-heating seasons (Fig. 7), and summary data on heating (Fig. 8) and non-heating seasons (Fig. 9).

The next stage is to analyze the spread diagrams. Fig. 10–15 show charts on the seasonal energy consumption during heating seasons (heat consumption and total energy consumption).
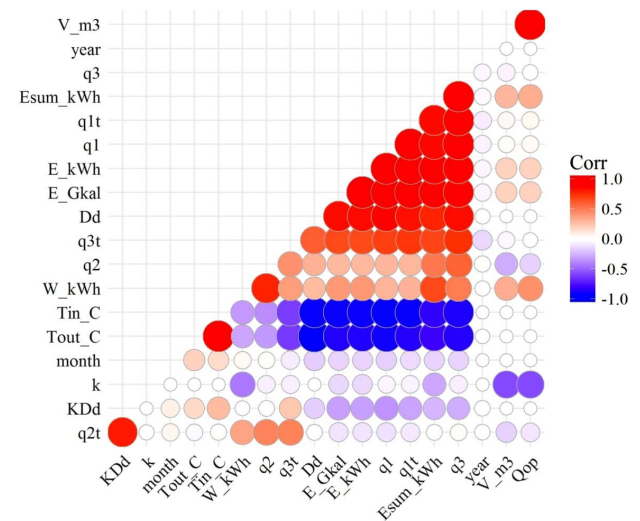

Fig. 5. Correlation matrix of monthly data (not split into heating and non-heating seasons) on energy consumption
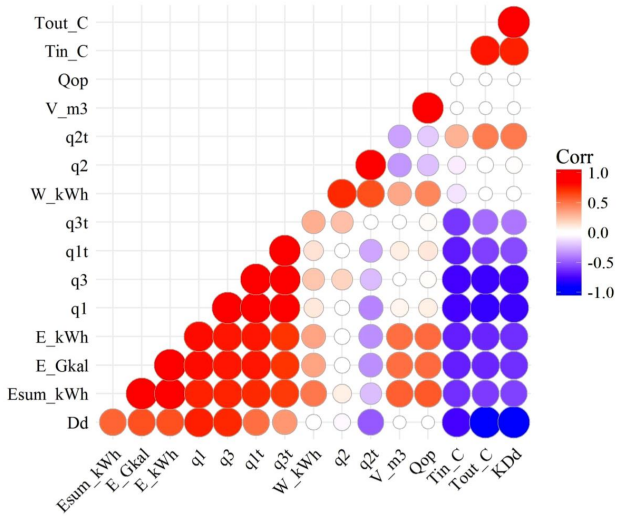
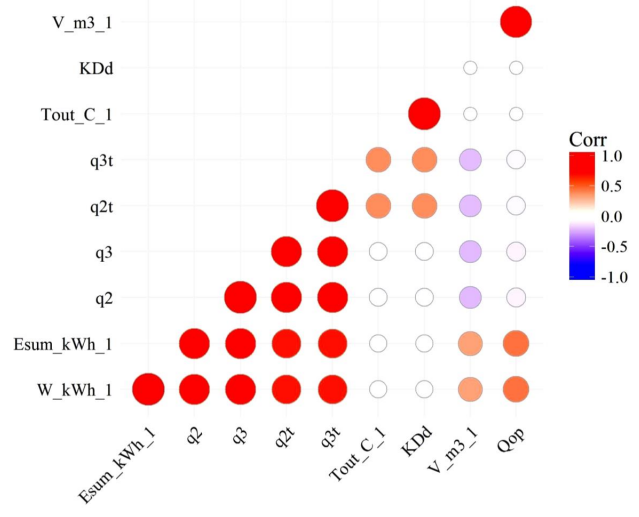**Fig. 6.** Correlation matrix of monthly data on energy consumption during heating seasons

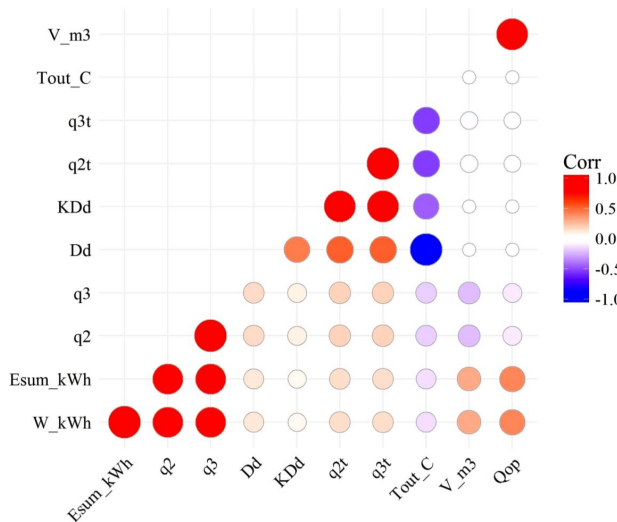**Fig. 9.** Correlation matrix of total data on energy consumption during non-heating seasons

Fig. 7. Correlation matrix of monthly data on energy consumption during non-heating seasons
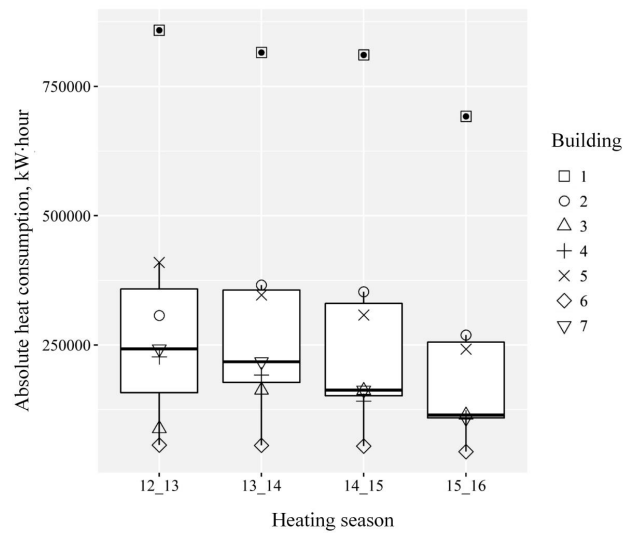

Fig. 10. Spread diagram of total seasonal data on the absolute heat consumption by buildings
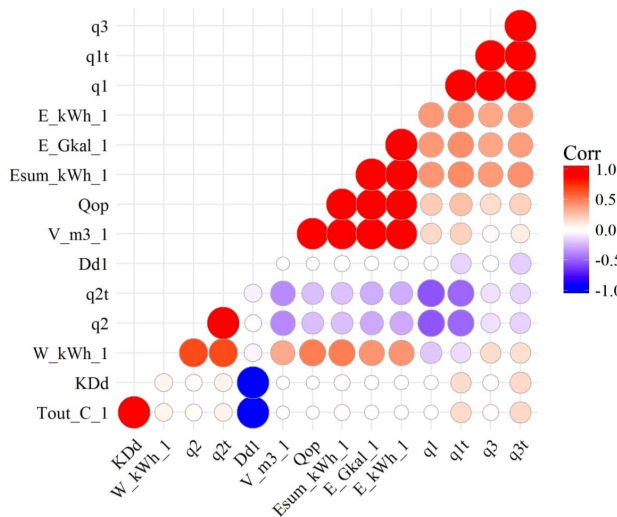

Fig. 8. Correlation matrix of total data on energy consumption during heating seasons
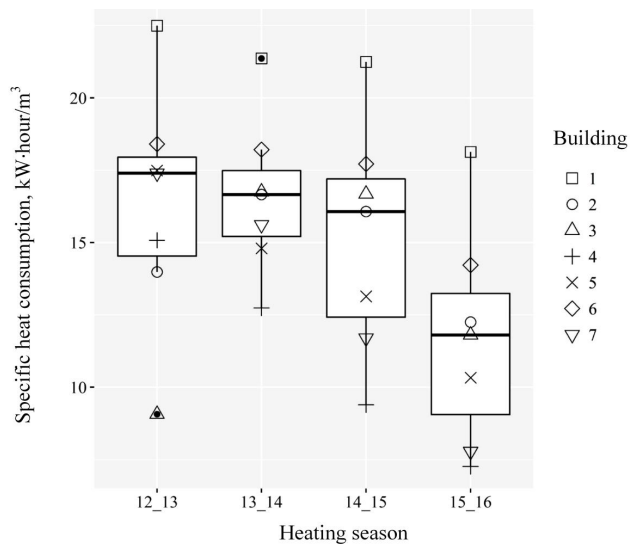

Fig. 11. Spread diagram of total seasonal data on specific heat consumption by buildings
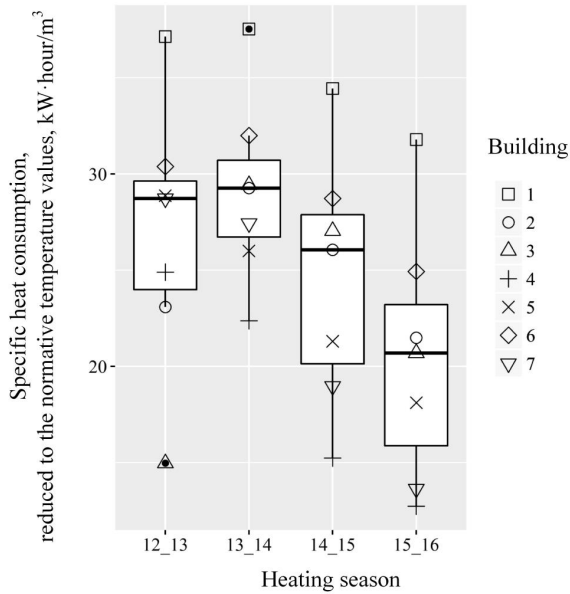
Fig. 12. Spread diagram of total seasonal data on specific heat consumption by buildings, reduced to the normative values for the outdoor and indoor air temperature
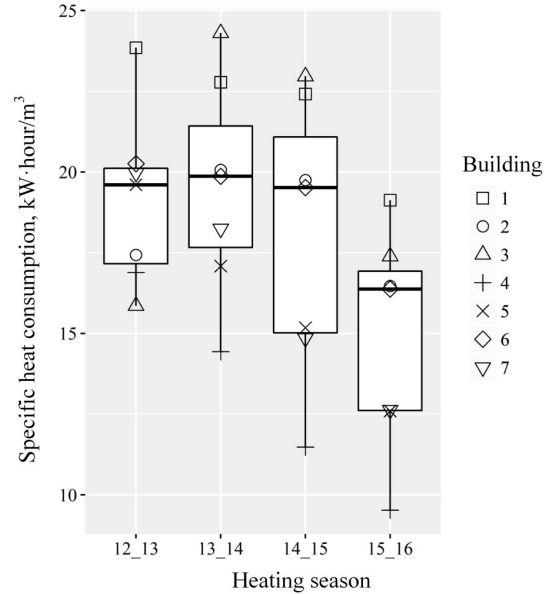

Fig. 14. Spread diagram of total seasonal data on the specific energy consumption by buildings
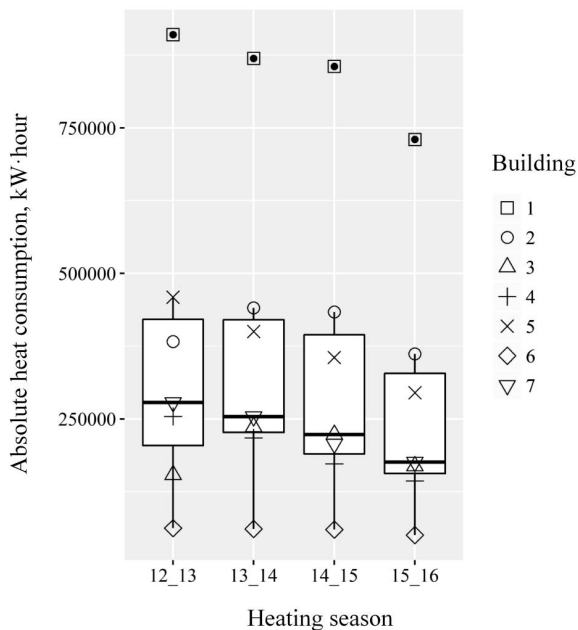

Fig. 13. Spread diagram of total seasonal data on the absolute energy consumption by buildings
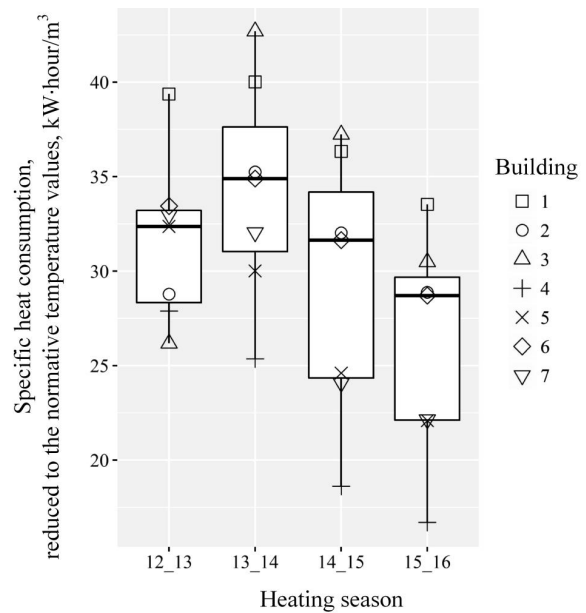

Fig. 15. Spread diagram of total seasonal data on the specific energy consumption by buildings, reduced to the normative values for outdoor and indoor air temperature

We analyzed data on heat consumption. Over a heating season of 2012–2013, the absolute heat consumption by building No. 5 (at thermal load that equals 0.2761 Gcal/h) exceeded the same indicator of building No. 2 (at thermal load that equals 0.3019 Gcal/h) by 102,568 kW·hour (Fig. 10).

Over the following seasons, building No. 2 exceeded indicators of building No. 5: during the heating season of 2014–2015 the absolute heat consumption by building No. 2 was 352,956 kW·hour (which made it possible to leave the third quartile – 330,374 kW·hour).

Specific heat consumption of building No. 3 during the heating season of 2012–2013 is minimal and is a kink, although analysis of the absolute heat consumption does not make it possible to determine it.

Specific heat consumption by building No. 4 over the heating season of 2012–2013 exceeds the same magnitude for building No. 2, given the fact that the heat load of building No. 2 is almost two-fold larger than the heat load of building No. 4 (Fig. 11).

Accounting of specific heat consumption, whose magnitude is reduced to the normative indoor and outdoor air temperature (Fig. 12), makes it possible to determine changes in the trend. Over the heating season of 2013–2014, at a maximal average temperature of the heating season (minimum number of degree-days), there is a substantial increase in specific heat consumption by buildings (a kink for building No. 1).

Specific heat consumption over the heating seasons of 2013–2014 and 2015–2016 with a close enough number of degree-days is significantly different.

Over the heating season of 2015–2016, there is a significant reduction in the absolute heat consumption compared with the season of 2013–2014, given the small difference of average air temperature over the heating seasons.

The analysis of the absolute total energy consumption over the heating periods from 2012 to 2016 allows us to argue about a trend towards lower energy consumption by buildings. Despite this, the value for the absolute total energy consumption of building No. 1 is a kink (Fig. 13).

Over the heating season of 2012–2013, consumption by building No. 5 exceeded the consumption by building No. 2, starting at the season of 2013–2014 energy consumption by building No. 2 exceeds that by building No. 5.

In the transition to specific total energy consumption, the trend is maintained, and the dependence of seasonal data on temperature manifests itself slightly.

We observe a significant jump in the value for specific energy consumption by building No. 3, which is observed somewhat weaker in the data on absolute total energy consumption (over the period from 2012 to 2014) (Fig. 14).

Over the heating seasons of 2013–2014 and 2014–2015, consumption of kW·hour per unit volume by the building No. 3 exceeded this indicator for building No. 1, which has the largest volume (Fig. 15).

5. 3. Development of a web-based solution to analyze heat consumption by apartments

To simplify the procedure for analysis of data on the heat consumption by apartments, we developed an appropriate web application. The proposed tool for implementing the client-server architecture to examine data on heat consumption is the framework Shiny, which makes it possible to rapidly and conveniently develop a web-interface using open-source libraries and the programming language R.

A user that has any device connected to the Internet can, by following the procedure of authorization or registration, get an opportunity and a toolset to process data on heat consumption by apartments. The data entered by the user are stored in a MySQL database and, given the existence of several types of users, the set of functions that can be performed by the user is limited. When trying to perform a function that is not included in the list provided for a given user, there is an error message.

Depending on the type of a building whose data on heat consumption will be examined, based on the constructed structures of information flows, various forms of data are implied.

We have implemented a possibility for the user to download a file with monthly data or to input them to the developed electronic form, as well as to convert them into values for seasonal consumption and to create additional parameters for comparison: specific consumption, energy efficiency coefficient, etc. Additionally, the software stores and displays in the additional chart values for the outdoor air temperature, obtained from the Internet resources. Parameters for setting the visualization make it possible to not only choose the Cartesian axes, but also perform additional splitting based on quality variables: the number of rooms, availability of a counter, etc.

Fig. 16 shows a fragment of the web-interface of the developed software that demonstrates seasonal indicators.

There is a possibility to additionally display the norms of consumption of heat energy for the system of centralized heating (heat supply). Solving the task on the generation and storing generated reports predetermines the necessity of ensuring printing quality in the assigned format. In this case, such reports would be easily understood by someone who is not necessarily a specialist in energy management. The implementation is based on using the language of data markup – Markdown.
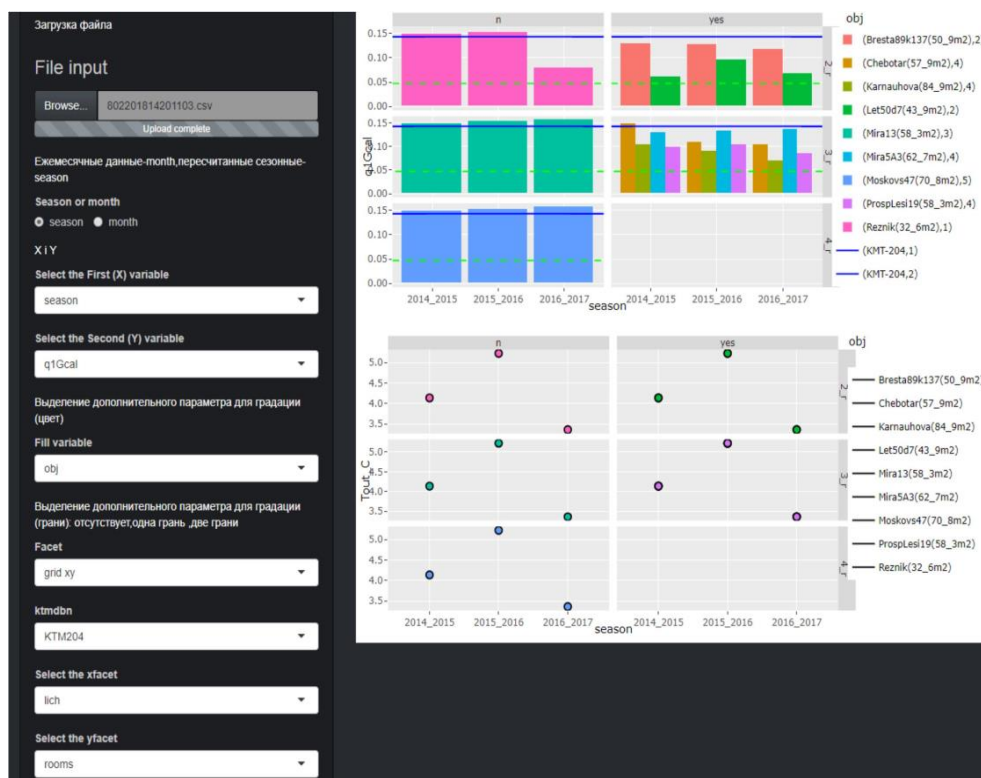

Fig. 16. Fragment of the web-interface of the developed software that demonstrates seasonal indicators

6. Discussion of results of studying the energy consumption by buildings

Comparing the correlation matrices for general data (Fig. 5), monthly data on heating (Fig. 6) and non-heating seasons (Fig. 7), and summary data on heating (Fig. 8) and non-heating seasons (Fig. 9), allows us to conclude:

1. The transition from general data to data on a heating period reveals a link between specific heat consumption and total specific energy consumption, which are reduced to the normative indoor and outdoor air temperatures.

2. The relationship between the absolute heat- and electricity consumption is expressed weakly and decreases in the transfer from general data to data on heating periods.

3. There is a link between absolute and specific heat consumption. Additional consideration of normative parameters for temperature slightly degrades this relationship, but additionally takes into consideration the volume and the ambient temperature.

4. Absolute heat consumption is strongly dependent on the volume of buildings, both for general data and the data on heating periods.

5. During the transition to summarized data on heating periods the relationship between specific heat- and energy consumption becomes more pronounced.

The analysis of spread diagrams for total seasonal data (Fig. 10–15) indicates the tendency to reduce energy consumption, due to the use of an automated system for monitoring and control over heat consumption at separate buildings. There is a significant heat consumption by the administrative building and over-expenditure of electricity by the building that hosts a diner.

The obtained results of an initial analysis could form the basis for constructing a model, which, by taking into consideration the structure of informative data and goals, would be able to solve the set analytical problems, to visualize results and to generate reports.

Our conclusions are relevant from a practical point of view, at they make it possible to reasonably approach processes of energy monitoring and energy management.

It should be noted that this paper does not give the algorithm to work with missing data, and there are no constraints on the size of the sample, which is dictated by the features in using the test to evaluate the normality of distribution: the number of values in the range from 3 to 5,000. In addition, there are certain limitations on the application of a given procedure in its original form, since part of the *a priori* data can be unavailable. Thus, the data on the amount of thermal load are derived from technical documentation for the object, which may complicate accessing them, especially in the case when energy consumption concerns not buildings as a whole but individual consumers. Since there is a possibility to generate and store standard reports on the original analysis of data, which are created using the technology of literate programming, there are several strict requirements to the examined data. Developing new attributes to compare energy consumption by buildings leads to an increase in the dimensionality of data, and, consequently, to an in increase in the time for various activities over data, which could be critical for larger samples over significant periods of time. The inability to remove the named limitations within the framework of this study gives rise to a potentially interesting area for further research.

6. Conclusions

To increase the efficiency of heat supply, it is necessary to devise information solutions aimed at intelligent processing of data on energy consumption and solving the tasks on reducing energy resources consumption.

1. In the course of our research, we designed the structure of information flows to acquire data from available sources. We have established the features of data depending on the type and functional purpose of buildings. Study into the structure of data has confirmed the need for the creation of additional attributes in order to analyze energy consumption by buildings in the form of specific heat consumption taking into consideration the temperatures of outdoor and indoor air, a class of energy efficiency, as well as tariffs per unit of a heated area.

2. Given the complex structure of data on the energy consumption by buildings, we have devised a procedure for initial analysis based on the sectoral methodology CRISP-DM. The approbation of the procedure, performed employing data on the academic buildings at KrNU allows us to assert that a given algorithm could be used for a preliminary analysis of data based on the Data Mining tools in the future.

3. To improve and facilitate the analysis of data on heat consumption by apartments, we have developed a web application. By using it we collect data and visualize them. There is a possibility to automatically recalculate existing monthly data into seasonal, to form new parameters for analysis, to download a calendar of temperature depending on geographic location of the user, etc. One can also compare heat consumption by apartments to the set requirements for the efficient use of energy resources in terms of the maximum specific need for heating.

References

1. A review of data-driven approaches for prediction and classification of building energy consumption / Wei Y., Zhang X., Shi Y., Xia L., Pan S., Wu J. et. al. // Renewable and Sustainable Energy Reviews. 2018. Vol. 82. P. 1027–1047. doi: https://doi.org/10.1016/j.rser.2017.09.108

2. Perekrest A. L., Romanenko S. S. Scientific and applied aspects saving energy and resources in the municipal energy sector // Elektromekhanichni i enerhozberihaiuchi systemy. 2015. Issue 2. P. 162–170.

3. Complex information and technical solutions for energy management of municipal energetics / Perekrest A., Shendryk V., Pijarski P., Parfenenko Y., Shendryk S. // Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017. 2017. doi: https://doi.org/10.1117/12.2280962

4. Balco P., Drahošová M., Kubičko P. Data analysis in process of energetics resource optimization // Procedia Computer Science. 2018. Vol. 130. P. 597–602. doi: https://doi.org/10.1016/j.procs.2018.04.109

5. Zakovorotnyi A., Seerig A. Building energy data analysis by clustering measured daily profiles // Energy Procedia. 2017. Vol. 122. P. 583–588. doi: https://doi.org/10.1016/j.egypro.2017.07.353

6. Arregi B., Garay R. Regression analysis of the energy consumption of tertiary buildings // Energy Procedia. 2017. Vol. 122. P. 9–14. doi: https://doi.org/10.1016/j.egypro.2017.07.290

7. Energy consumption forecasting based on Elman neural networks with evolutive optimization / Ruiz L. G. B., Rueda R., Cuéllar M. P., Pegalajar M. C. // Expert Systems with Applications. 2018. Vol. 92. P. 380–389. doi: https://doi.org/10.1016/j.eswa.2017.09.059

8. Ahmad M. W., Mourshed M., Rezgui Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption // Energy and Buildings. 2017. Vol. 147. P. 77–89. doi: https://doi.org/10.1016/j.enbuild.2017.04.038

9. Jeffrey Kuo C.-F., Lin C.-H., Lee M.-H. Analyze the energy consumption characteristics and affecting factors of Taiwan's convenience stores-using the big data mining approach // Energy and Buildings. 2018. Vol. 168. P. 120–136. doi: https://doi.org/10.1016/j.enbuild.2018.03.021

10. Parfenenko Yu. V., Shendryk V. V., Galichenko O. S. Prediction the heat consumption of social and public sector buildings using neural networks // Radioelektronika, informatyka, upravlinnia. 2015. Issue 2. P. 41–46. URL: http://nbuv.gov.ua/UJRN/riu_2015_2_7

11. Slabchenko O., Sydorenko V., Siebert X. Development of models for imputation of data from social networks on the basis of an extended matrix of attributes // Eastern-European Journal of Enterprise Technologies. 2016. Vol. 4, Issue 2 (82). P. 24–34. doi: https://doi.org/10.15587/1729-4061.2016.74871

12. Building energy performance diagnosis using energy bills and weather data / Geng Y., Ji W., Lin B., Hong J., Zhu Y. // Energy and Buildings. 2018. Vol. 172. P. 181–191. doi: https://doi.org/10.1016/j.enbuild.2018.04.047

13. A study and a directory of energy consumption data sets of buildings / Babaei T., Abdi H., Lim C. P., Nahavandi S. // Energy and Buildings. 2015. Vol. 94. P. 91–99. doi: https://doi.org/10.1016/j.enbuild.2015.02.043

14. Perekrest A. L., Zagirnyak M. V. Opyt vnedreniya i ispol'zovaniya avtomatizirovannoy sistemy monitoringa temperaturnyh rezhimov i udalennogo upravleniya teplopotrebleniem Kremenchugskogo nacional'nogo universiteta // Elektrotekhnicheskie i komp'yuternye sistemy. 2014. Issue 15 (91). P. 423–426.