ECONTECHMOD. AN INTERNATIONAL QUARTERLY JOURNAL – 2018, Vol. 07, No. 2, 45-52

Energy consumption monitoring system and energy data processing tools

I. Tsmots, Yu. Tsymbal, V. Roman, D. Karpa

Lviv Polytechnic National University, Lviv, Ukraine; e-mail: ivan.tsmots@gmail.com

Received April 19.2018: accepted June 05.2018

Abstract. It is determined that monitoring system of energy consumption at the enterprises in Lviv region should provide the implementation of the following tasks: energy data collection on a real-time basis; integration and processing of information resources justifying energy consumption at both enterprise and regional levels; communication between regional authorities and entrepreneurs; quick access of regional authorities to the latest information on energy consumption; analysis of energy consumption dynamics, structure and industrial energy efficiency; short-term and medium-term predictions of energy consumption and energy efficiency of regional economy; daily energy audit of enterprises of the region; rapid identification and response to energy loss (accidents, damage, illegal use etc); working out the effective managerial decisions based on processing of large amount of information related to energy consumption. The structure of the system of energy consumption monitoring in Lviv region was developed. It consists of two subsystems - information and analytical ones. The technical basis of the information subsystem comprises wireless sensor networks ensuring energy data collection on a real-time basis. Analytical subsystem consists of energy data warehouse, ETL, OLAP, BI, Data Mining tools, prediction, visualization and decision support tools. The components of the analytical subsystem providing for visualization and evaluation of industrial energy consumption were developed.

Keywords: monitoring system, information subsystem, analytical subsystem, sensor networks, visualization, energy data, components.

PROBLEM STATEMENT

The monitoring system of energy consumed by the enterprises in Lviv region is an information system that provides for the energy audit (electricity, water, heat, gas) and plans activities aimed at energy conservation. The monitoring system of energy consumption at the enterprises in Lviv region should involve a consolidated collection, storage, processing, and analysis of information on industrial energy consumption as well as working out the effective managerial decisions to improve their energy efficiency. The characteristic features of the monitoring system of energy consumption at the enterprises in Lviv region should include the following:

- a comprehensive monitoring of energy consumption at the enterprises (institutions) on a real-time basis;
- integration and processing of information resources justifying energy consumption at both enterprise and regional levels;
- communication between regional authorities and entrepreneurs;
- quick access of regional authorities to the latest information on energy consumption at the regional enterprises (institutions);
- analysis of energy consumption dynamics, structure, and energy efficiency at the regional enterprises (institutions);
- short-term and medium-term predictions of energy consumption and energy efficiency of the regional economy;
- the daily energy audit of enterprises (institutions) of the region;
- rapid identification and response to energy loss (accidents, damage, illegal use etc);
- working out the effective managerial decisions based on the processing of a large amount of information related to energy consumption.

Tools for processing of data presented by the monitoring system of energy consumption at the enterprises of Lviv region should provide for graphical representation of information on energy consumption with the help of graphs and charts. In addition, such tools should generate the analytical reports that assess the energy consumption dynamics according to the types of recourses and conduct a comparative analysis of industrial energy consumption.

When developing the energy consumption monitoring system for Lviv region one should take into consideration an extensive use of TV and web technologies, databases, data acquisition, processing, visualization tools as well as tools for making managerial decisions. Managerial decisions should focus on optimization of the energy consumption pattern, reduction of the energy used per production unit, work performed and services rendered, reduction of non-productive losses of fuel and energy recourses.

Therefore, the improvement of the energy efficiency at the enterprises of the region with the use of energy consumption monitoring system is of particular relevance.

ANALYSIS OF RECENT RESEARCHES AND PUBLICATIONS

In recent years, there have been a lot of research and publications focusing on the use of energy consumption monitoring system at the municipal, district and regional levels [1-23]. Let us discuss the most important publications.

The papers [1-15] present the analysis of energy consumption monitoring systems and multi-level energy efficiency management systems used at the individual as well as regional enterprises. The drawbacks of such systems are the absence of the separate database on each control layer as well the fact that system structure does not involve the use of the modern Internet technologies.

The papers [1-8] focus on the issues of the energy management research as an effective means for increasing energy efficiency of the enterprises and regional economy. However, tools for energy consumption processing and visualization have not been fully developed by now.

The analysis of publications [9-15] justifies that working out the effective managerial decisions within the scope of energy consumption monitoring systems requires the development of tools for analytical and intelligent processing of energy consumption indicators.

The papers [2-7] deal with technical and engineering aspects of the development of energy consumption monitoring systems; the modern technologies of the collection, storage, and processing of information are reviewed and their usefulness for such systems is proven. However, the papers under analysis do not focus on the comprehensive approach to the development of such tools and their adapting to energy efficiency management problems at the individual as well as regional enterprises.

The analysis justifies that the development of the monitoring system of energy consumption in the region requires a comprehensive approach involving modern telecommunications and web technologies, databases, tools for collection, processing, and visualization of energy consumption indicators.

AIM AND OBJECTIVES OF THE RESEARCH

The aim of the research is to develop a structure of the regional energy consumption monitoring system and its components.

To achieve the aim the following objectives should be accomplished:

- to develop the structure of the regional energy consumption monitoring system;
- to analyze and choose means of implementation of the monitoring system components;
- to develop the components of the energy consumption monitoring system.

KEY FINDINGS

Structure of the regional energy consumption monitoring system. Structure of the monitoring system of energy consumption in Lviv region is focused on using telecommunications and web technologies, databases, tools for collection, analytical and intelligent processing of energy data, visualization of the results of processing and making managerial decisions. Energy consumption monitoring system consists of two subsystems: information and analytical ones. The structure of the developed energy consumption monitoring system is presented in Fig. 1, where TCP/IP - Communication Protocol Suite, ETL - Extract, Transform and Load, OLAP - Online Analytical Processing; BI - Business Intelligence, *N* - the number of sensor system nodes.

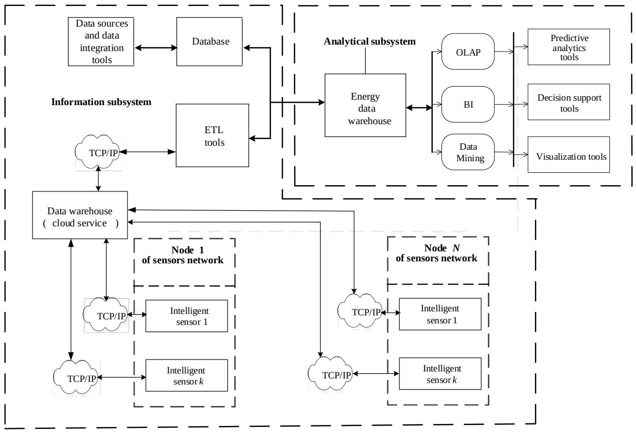

Fig. 1. Structure of the regional energy consumption system

The technical basis of the information subsystem comprises the wireless sensor networks characterized by self-organization and adaptability to the changes in environment and infrastructure. In addition, wireless node hardware and their networking protocols are optimized with the respect to energy consumption. Generally, the wireless sensor network is a distributed system consisting of the small intelligent (though not always) sensory units linked to the cloud data warehouse. The sensor network may include the following components: distributed sensors, meter readers, computation, and transmission units. All intelligent sensors interact through the global or local infrastructure for the information exchange of access to the cloud data warehouse. The status of sensors and the number of connections can vary over time. Additionally, the collection and maintenance of energy data may be done manually by means of recording in the database.

The cloud data warehouse provides open interfaces to access and publish integrated data controlled by the cloud. Any Internet user can access these data through public, documented access interfaces provided by the cloud. Enterprises and institutions are responsible for data collection in the cloud. Energy data from enterprises are recorded in the cloud data warehouse with the help of HTTP-based interfaces. To work with cloud data warehouse one is recommended to use the interface with HTTP REST protocol - a simple and unified interface for network communication. Such interface provides an opportunity to publish or read the data from the service directly as well as make requests for regulating actions. The example of such actions can be a request for service's installation of data streaming read based on TCP sockets.

Service can provide file interfaces to publish data files. Such interfaces can be FTP services, local file systems or cloud file systems, e. g. AWS S3. Data can be stored in popular clouds, such as AWS, Azure and Google Cloud.

Analytical subsystem consists of energy data warehouse, ETL, OLAP, BI, Data Mining tools, prediction, visualization and decision support tools. Energy data warehouse is based on big data technologies allowing effectively storing, processing and providing access to large volumes of data in the distributed mode. Moreover, such technologies provide for the effective processing of unstructured data. One should use OLAP servers to access OLAP technologies in the data warehouse. Users can access these tools either directly via interfaces or via software applications (desktop, web or mobile ones).

Tools for monitoring system components implementation. The following technical tools are chosen for implementation of the energy consumption monitoring system components:

- data extraction - web spider created with the help of Scrapy framework that provides a flexible infrastructure for web spider creation;
- operational databases - NoSQL database (MongoDB) that ensures quick work with individual documents and support the effective scaling in;
- the data bus is represented by Kafka message broker that is able to manage large volumes of messages, has broad means of integration with various components of data recording and reading, and supports clustering;
- data is processed with the help of Spark framework based on Hadoop YARN resource manager;
- thread-specific data is processed with Spark Streaming that ensures processing of new messages received by the message broker while reducing the time recent data are available in the data warehouse to a few seconds;
- the data warehouse is based on Hadoop HDFS distributed file system;
- batch processing is carried out with the help of Spark Core and Spark SQL components based on Hadoop YARN;
- tabular data access based on SQL is organized with the help of HIVE on the basis of Spark SQL engine;
- distributed intelligent data processing is performed with the use of TensorFlow;
- Druid OLAP server providing for operation of OLAP technology on a real-time basis and with large volumes of data;
- server integration with Kafka broker;
- the following interfaces are available for users: Hue for creating SQL HIVE base queries; Pentaho Report Designer for creating reports and graphs based on HIVE; a website providing access to graphic data of intelligence and OLAP data processing;
- Cassandra database for storing data processing results and MySQL for storing component metadata;
- Zookeeper for service synchronization and availability;
- Docker containerization technology and Kubernetes container orchestration technology.

Tools used for monitoring system implementation are summarized in Fig. 2.

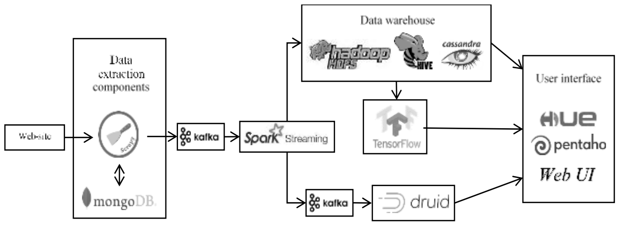

Fig.2 Tools for monitoring system implementation

Data extraction component is a web spider based on Scrapy technology and online NoSQL database MongoDB. The module will be written in Python. The complexity of collecting data from websites is explained by the lack of information on updates of the certain website data. The collection component should occasionally review all data available on the website and identify the subsequent data updates on the basis of the state stored in the online database. Data updates are sent to the message bus for further processing and storage in the warehouse. Data extraction component operates in two modes:

- package mode: all the data on the buildings indicated on the website are collected at one time;
- incremental update: data updates are collected.

Components of the analytical subsystem. Data is preprocessed and stored in the warehouse in the streaming mode providing for a minimum delay between data entering and storing in the data warehouse. Spark Streaming is used for stream data processing. The technology undertakes micro-batch stream processing and provides for the following functions necessary for data processing:

- window procedure and sliding window processing;
- stateful batch processing;
- batch processing of structured data using Spark SQL Datasets;
- state, settings and data are saved to the reliable intermediate storage to be restored in case of hardware or process failure;
- Exactly once delivery semantics for working with Kafka broker.

Preprocessing components can operate in both distributed mode and parallel mode with the application of Kafka broker. The number of Spark Streaming processes should be equal to or less than the number of Kafka topic partitions for low-load computing. Heavy computing may involve more processes.

Spark Streaming operated on the basis of Hadoop YARN resource manager. The resource manager is responsible for the number of processes and their management, ensures availability and restoration of the processes after hardware or software failure. Data preprocessing involves the following steps:

- Verifying data formats and their compliance with the requirements. Data, which failed to be verified, are filtered out and recorded to a special area with invalid data for further analysis.
- Noise reduction.
- Converting data to a specific format and saving to the following repositories: Kafka topic for Druid server; staging area in HDFS file system.
- Counting statistical data, such as the number of processed and filtered messages.

The general scheme of data processing in the data warehouse is presented in Fig. 3.

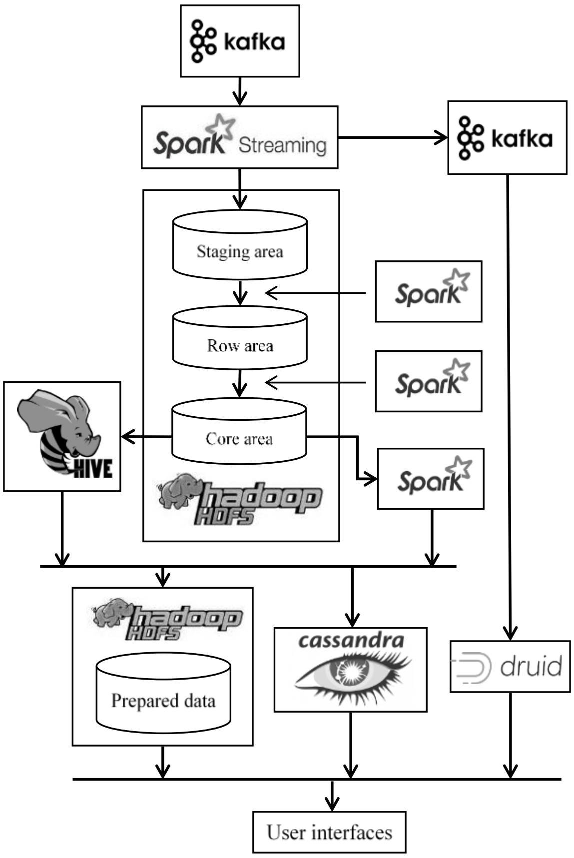

Fig. 3. Scheme of data processing in data warehouse

Files with new data are stored in the staging area. Spark Batch process runs periodically and aggregates them into large files and saves to the raw area. Then there is a start of the process processing new data in the raw area and saving them to the core area. This process provides for additional data conversion and enrichment. Based on the building address, geographic coordinates (latitude and longitude) are calculated and added. Data are normalized to the third normal form, the building address is subdivided into the separate sections indicating country, city, street and building number.

One can access raw and prepared data by means of SQL requests using HIVE technology. Data are prepared for user interface by means of HIVE requests and Spark processes. The results are saved to HIVE table in HDFS system (Prepared data) or Cassandra database. Additionally, Spark Streaming process saves data to Kafka topic for it to be uploaded to Druid OLAP server. The server supports the integration of data from Kafka broker on a real-time basis.

Meter readings of various buildings are stored in the relational database, which structure is presented in Fig. 4.

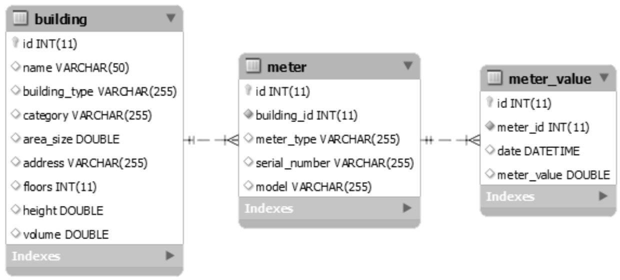

Fig. 4. Raw database structure

The database consists of three tables: building - building data; meter - meters in the building; meter_value - meter values in a certain period of time. Raw area preserves extracted data unaltered. That is why its structure is similar to the one presented in Figure 4. Core area contains data normalized to the third normal form, cleansed and supplemented by additional fields (geographical coordinates of the building: longitude and latitude). These data are intended for working with user interfaces. The cleansed database structure is presented in Fig. 5.

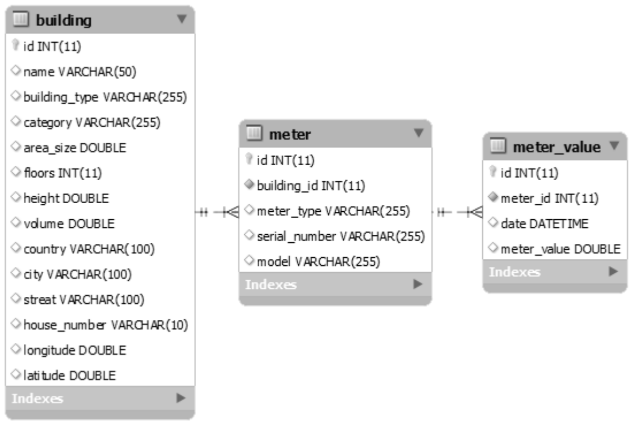

Fig. 5. Cleansed database structure

Data extraction component is a web spider based on Scrapy technology and online NoSQL database MongoDB. MongoDB is used to save the last data status extracted by the web spider - the procedure that allows one to monitor data changes in the sources, particularly new meter readings. Information about building, its meters, meter values are saved to a single document for consistency.

In case of OLAP analysis, cleansed data (core area) are presented in the form of a star schema or denormalized to a single table. Different OLAP servers use the different structure of data to work with. For example, Druid OLAP server handles denormalized data.

The development of regional energy consumption map. Energy consumption map is prepared with the use of SQL script presented in Fig. 6. The sample of energy consumption data per week (from 23 September 2017 till

30 September 2017) was collected and geographical center of consumers was specified.

```
DROP TEMPORARY TABLE IF EXISTS subset_in_time_range;
CREATE TEMPORARY TABLE IF NOT EXISTS subset_in_time_range AS (
    SELECT address, MAX(meter_value) - MIN(meter_value) as meter_delta_value,
        formated_date, latitude, longitude, building_id, meter_id
    FROM (
        SELECT b.address, v.meter_value, STR_TO_DATE(`date`, '%e %M %Y, %H:%i') as formated_date,
            mapping.latitude, mapping.longitude, b.id as building_id, m.id as meter_id
        FROM meter_value as v JOIN meter as m
        ON v.meter_id = m.id
        JOIN building as b
        ON b.id = m.building_id
        JOIN address_geolocation_mapping as mapping
        ON mapping.address = b.address
        WHERE v.meter_value > 0 AND m.resource_type = 'електроенергія'
    ) as sub

    WHERE formated_date between '2017-09-23 00:00:00' and '2017-09-30 23:59:00'
    GROUP BY meter_id, building_id
);

SELECT 'address', 'meters_value', 'date', 'latitude', 'longitude'
UNION ALL
SELECT address, SUM(meter_delta_value) as meters_value, formated_date, latitude, longitude
FROM subset_in_time_range
GROUP BY building_id
```

Fig. 6. Data preparation script of energy consumption map

The example of industrial energy consumption per week with geographical coordinates sorted in descending order is presented in Table 1.

Table 1.
Industrial energy consumption per week with geographical coordinates

Addres	Total consumption (kW)	Latitude	Longitude
Жовква, вул Коновальця,3	873 941,00	50,05722	23,976
Львів, вул, Пилипа Орлика,6	746 901,33	49,35341	23,511
Львів, вул. Чернігівська 7	709 718,33	49,83921	24,046

Using data on industrial energy consumption per week and Google Maps API libraries, the energy consumption map presented in Fig. 7 was designed.

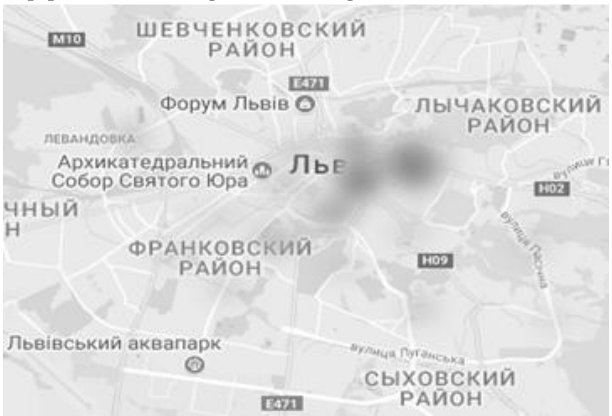

Fig 7. Energy consumption map (per week)

Energy consumption intensity in different parts of the region is marked with a blue color. The darker the color, the more electricity was consumed at this location per week. If there is the blue color on the map, there is the data on the energy consumed at the specific places on the map. The energy consumption map allows one to see the areas with the highest or the lowest energy consumption rate.

Energy Consumption Visualization. The data preparation script displaying the electricity consumed by the enterprise for a certain period of time with different time intervals was developed. The example of the graph showing energy consumed by the enterprise with a daily interval is presented in Fig. 8.

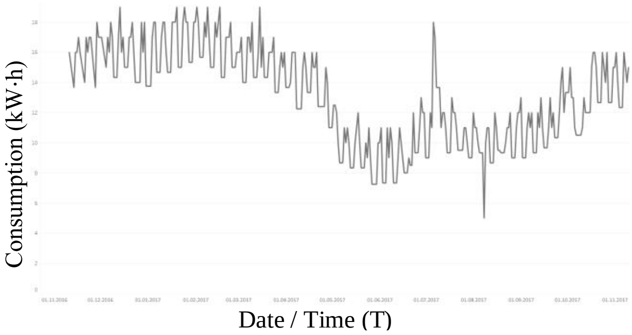

Fig. 8. The graph for energy consumed by the enterprise with a daily interval

This graph shows the daily energy consumption rate during one month. Additionally, the managerial decisions involve graphs for energy consumption with an hour, week, and month interval.

Comparison of energy efficiency of the enterprises. To compare the energy efficiency of various enterprises, the data should be prepared in the following way:

- energy consumption rate of each enterprise is identified by summing up all meter readings;
- the energy efficiency ratio (showing the amount of consumed electricity per unit area) should be determined as follows: <C = E / S>, where C - efficiency ratio, E - consumed electricity, S - the building area;
- the energy efficiency ratio of the enterprises should be normalized: <Cn = C/max(C)>, where Cn - a normalized energy efficiency ratio, max (C) - maximum ratio value among all enterprises.

Control parameter lies within the range from 0 to 1, , C_n is [0.0, 1.0]. According to this parameter, buildings are divided into 10 energy efficiency categories. The most energy efficient buildings (with regard to energy consumption per unit area) are included in the category 0, the least efficient ones - in the category 9. The buildings with efficiency ratio C_n = [0.0, 0.05] are included in the category 0, the ones with efficiency ratio C_n = [0.05, 0.15] - in the category 1 etc.

The script for weekly efficiency ratio of each enterprise was developed. Based on data on energy consumption and building area one can assess their energy efficiency. The area and electricity consumed by each enterprise per week are presented in Figure 9.

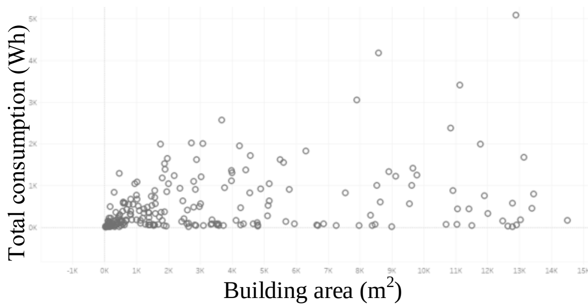

Fig. 9. Diagram of buildings grouped by the area and consumed electricity

Diagram of buildings grouped by the area and consumed electricity allows one to identify trends in the energy efficiency of the buildings. This diagram helps identify buildings dissenting from the general trend. The additional building characteristics and forms of energy may be used to assess energy efficiency rates more precisely.

Based on the results of building energy efficiency assessment the diagram of buildings grouped by energy efficiency categories was developed (Fig. 10).

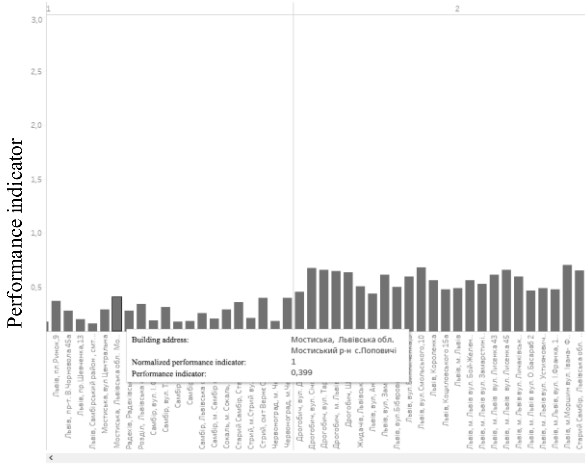

Fig. 10. Diagram of buildings grouped by energy efficiency categories

There are two categories of buildings: ones with a low energy efficiency ratio are on the left, ones with a high energy efficiency ratio are on the right.

The block graph representing buildings grouped by energy efficiency categories is presented in Fig. 11.

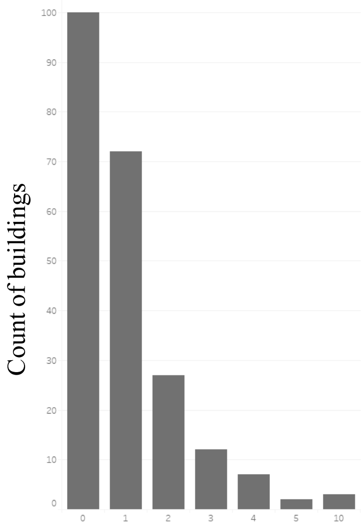

Fig. 11. Graph for buildings grouped by energy efficiency categories

Each energy efficiency category shows the ratio of consumed energy to the building area. Category 0 is considered to be the most energy efficient, category 9 - the least energy efficient.

CONCLUSIONS

1. It was determined that monitoring system of energy consumption at the enterprises in Lviv region should provide for implementation of the following tasks: energy data collection on a real-time basis; integration and processing of information resources justifying energy consumption at both enterprise and regional levels; communication between regional authorities and entrepreneurs; quick access of regional authorities to the latest information on energy consumption; analysis of energy consumption dynamics, structure and industrial energy efficiency; short-term and medium-term predictions of energy consumption and energy efficiency of regional economy; daily energy audit of enterprises of the region; rapid identification and response to energy loss (accidents, damage, illegal use etc); working out the effective managerial decisions based on processing of large amount of information related to energy consumption.

2. Structure of the monitoring system of energy consumption in Lviv region was developed; it consists of information and analytical subsystems focused on using telecommunications and web technologies, databases, tools for collection, analytical and intelligent processing of energy data, visualization of the results of processing and making managerial decisions.

3. The technical basis of information subsystem of the energy consumption monitoring system in Lviv region comprises wireless sensor networks focused on energy data collection on a real-time basis.

4. Components of the energy consumption monitoring system in Lviv region were developed; they provide for visualization and assessment of energy consumed by the enterprise.

REFERENCES

1. **Rozen V.P., Chernyavskiy A.V., Yachnik E.A., Litvin V.I. 2009** Energy saving programs monitoring in public sector institutions // Visnyk KDPU of Mykhailo Ostrohradskyi. Edition 3 (56). Part 2. P.190-194.
2. **Teslyuk T., Tsmots I., Teslyuk V., Medykovskyy M., Opotyak Y. 2018** Architecture and Models for System-Level Computer-Aided Design of the Management System of Energy Efficiency of Technological Processes at the Enterprise // Advances in Intelligent Systems and Computing II, Advances in Intelligent Systems and Computing 689. Springer International Publishing AG. P.538-557.
3. **Medykovskyi M.O., Tsmots I.G., Skorokhoda. O.V.** Spectrum neural network filtration technology for improving the forecast accuracy of dynamic processes in economics // Actual Problems of Economics. №12 (162) 2014. Pp. 410-416.
4. **Medykovskyi M.O., Tsmots I.G., Skorokhoda O.V., Teslyuk T.V. 2016** Design of Intelligent Component

of Hierarchical Control System // Econtechmod. An International Quarterly Journal –Vol.5. No2.3-P.3-10.

5. **Medykovskyi M.O., Tsmots I.G., Tsymbal Y.V.** Intelligent data processing tools in the systems of energy efficiency management for regional economy // Actual Problems of Economics. №12 (150). P. 271-277.
6. **Tsmots I.G., Tsymbal Y.V., Tsmots O.I. 2012** Early-warning systems for enterprises using neural networks tools // Actual Problems of Economics. №10 (136). P. 283-291.
7. Intelligent Components of Integrated Automated Control Systems: monograph / **M.O. Medykovskyi, R.O. Tkachenko, I.H. Tsmots etc.** – Lviv: Publishing House of Lviv Polytechnic, 2015. – 315 p.
8. **Teslyuk T., Tsmots I., Teslyuk V., Medykovskyi M., Opotyak Y. 2017** Architecture of the management system of energy efficiency of technological processes at the enterprise // Proceedings of XIIth International Scientific and Technical Conference CSIT 2017. 5-8 September. Lviv, Ukraine – P.429-433.

Medykovskyi M.O., Tsmots I.H. Tsymbal Y.V. 2016 Information analytical system for energy efficiency management at enterprises in the city of Lviv (Ukraine) // Actual Problems of Economics. №1 (175). P.379-384.

10. **Medykovskyi M.O., Tsmots I.H., Podolskyi M.R. 2013** Substantiation of development principles and generalized structure construction of the informational analytical system for estimating, forecasting and managing energy efficiency of the region's economy // Visnyk NU "Lviv Polytechnic" "Computer Science and Information Technology", № 751. – Lviv. – P. 40-51.
11. **Tsmots I.H., Tesliuk V.M, Tesliuk T.V. 2017** Determining model for the investment priority of energy saving projects implementation on enterprise // Actual Problems of Economics. №1 (187). P. 398-408.
12. **Tsmots I.H., Tesliuk T.V., Opotiak Y.V. Tesliuk V.M. 2017** Architecture of the multilevel energy efficiency management system of the region // Visnyk NU "Lviv Polytechnic" "Computer Science and Information Technology" №864 Lviv. P.201-209.
13. **Tsmots I.H., Skorokhoda O.V., Roman V.I. 2016** Multilevel energy efficiency management systems data warehouses // Collection of scientific works "Modeling and Information Technology". Institute of Modeling Problems in Power Engineering. Edition 77. P. 192 – 197.
14. **Arakelov, V.E., Kremer, A.I. 1990** Methodological issues of energy saving. M: Energoatomizdat. 188 p.
15. **Ganzha, V.L. 2007** Basics of effective use of energy resources. Mn.: Beloruskaya nauka. 452 p.
16. Agile Best Practices for Data Warehousing [Electronic resource]. - Access mode: http://www.agiledata.org/essays/dataWarehousingBestPractices.html
17. Data Warehouse Design [Electronic resource]. - Access mode: http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300
18. Big Data Implementation vs. Data Warehousing [Electronic resource]. - Access mode: http://www.b-eye-network.com/view/17017
19. Lambda Architecture [Electronic resource] - Access mode: http://lambda-architecture.net/
20. Big Data and its Impact on Data Warehousing [Electronic resource] - Access mode: http://cdn.ttgtmedia.com/BeyeNETWORK/downloads/BigDataE-Book_final.pdf
21. Real-Time Stream Processing [Electronic resource] - Access mode: https://www.infoq.com/articles/stream-processing-hadoop
22. Batch vs Streaming Processing [Electronic resource] - Access mode: http://blog.syncsort.com/2017/07/big-data/big-data-101-batch-stream-processing/
23. Spark Streaming [Electronic resource] - Access mode: https://spark.apache.org/streaming/