

Critical Infrastructures in the Digital World

IWCI PROCEEDINGS OF THE 2026 INTERNATIONAL WORKSHOP

March 12 – 18, 2026
Bolshoe Goloustnoe

WORKSHOP PROCEEDINGS

IMPLEMENTATION OF A NEURO-ASSISTANT BASED ON THE RAG-SYSTEM

Alekseev Roman[1], Lukyanov Nikita[2], Massel Liudmila[1]

[1]Melentiev Energy Systems Institute SB RAS
[2]Irkutsk National Research Technical University
Irkutsk, Russia
alekseevrv@ex.istu.edu

The advancement of artificial intelligence systems has necessitated the creation of specialized neural assistants capable of managing not only data sets but entire knowledge ecosystems. At this stage of the research, the primary objective is the practical implementation and experimental verification of the selected design solutions. The aim of this study is to evaluate the chosen vector search methods, specifically by investigating the effectiveness of various vector data representation models and conducting a comparative analysis of distance metrics for determining semantic similarity in the context of neural assistant operation.

The primary element of the developed neural assistant is a system founded on the Retrieval-Augmented Generation (RAG) architecture. The selection of this paradigm arises from a significant limitation inherent in traditional generative models: neural assistants that function without external context depend solely on parametric memory, specifically, the knowledge the model has acquired during its training phase. This methodology presents several disadvantages: firstly, the model is susceptible to "hallucinations" when responding to inquiries that exceed its training data; secondly, it lacks the capability to rapidly adapt to alterations in the current information environment [1]. A critical component of the RAG neural assistant currently under development is the storage of vector representations, which facilitates rapid and precise access to pertinent context. The Qdrant vector database has been selected as the primary platform for this initiative. This selection is predicated upon a variety of architectural and operational advantages that are imperative for constructing a fault-tolerant knowledge ecosystem [2]. The principal distinction between Qdrant and "pure" vector indexes lies in its capability to store and filter metadata (payload). Each vector can encompass structured information, including document source, indexing date, content type, and access level, thereby enabling preliminary filtering of the search space and enhancing the relevance of the selection process. A notable feature of this project is the implementation of approximate nearest neighbor search (ANN) algorithms, such as graph-based indexing (HNSW), which contribute to the improvement of search performance. An essential practical objective is to optimize indexes to achieve an equilibrium between search accuracy and the consumption of computing resources. The principal focus of the experimental component was the examination of similarity metrics during the execution of search queries. Cosine distance (Cosine Similarity) was selected as the standard metric, as it allows for the measurement of the angle between vectors while disregarding their magnitude.

Thus, the developed tools lay the foundation for creating a fully-fledged neural assistant. Further research will focus on refining and implementing the developed mechanisms.

References

1. Melnikov A.V., Nikolaev I.E., Rusanov M.A., Abbazov V.R. Comparative analysis of RAG methods for building Russian-speaking intelligent services. Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics. 2025. Vol. 25. – No. 2. – P. 5–18. DOI: 10.14529
2. Obolensky D. M., Shevchenko V. I. Application of Large Language Models and the RAG in Intelligent Educational Ecosystems. Economics. Information technologies. 2024. Vol. 51. – No. 3. – P. 699–709. DOI 10.52575/2687-0932-2024-51-3-699-709.

ALGORITHM FOR MODELING A MULTI-STAGE MULTI-LEVEL PROBLEM OF OPTIMIZING THE PRODUCTION AND PROCESSING OF AGRICULTURAL PRODUCTS

Ananyev Leonid, Ivanyo Yaroslav

Irkutsk State Agrarian University named after A.A. Ezhevsky
Molodezhny, Russia
rector@igsha.ru

Optimizing the production and processing of agricultural products at an enterprise is a complex management task aimed at increasing the efficiency of resource use and improving planning quality. For enterprises that combine raw material production and processing, it is advisable to consider the processes in an interconnected manner: the distribution of acreage between crops and the technologies used determine the volume and quality of raw materials; the distribution of raw materials by use determines the utilisation of processing capacities; and the range of processed products influences the amount of added value and the final profit.

A significant factor is the uncertainty of operating conditions associated with changes in yields, selling prices, and resource availability. Therefore, when planning, it is advisable to use a multi-stage, multi-level model that allows for the development of optimal plans for various scenarios (unfavorable, average, and favorable). It is proposed to use a multi-level parametric programming model as a model for the production and processing of agricultural products.

The stage-by-stage structure of the problem is determined by the sequence of product flows within the enterprise. The first stage involves developing an optimal plan for the production of agricultural raw materials (selecting crops and output volumes) under resource and space constraints and meeting minimum production volume requirements for individual product types. Multi-level production optimization involves three process scenarios depending on external and production-economic conditions: average, favorable, and unfavorable.

The second stage involves determining a plan for processing the produced raw materials into final products, taking into account technological standards and production capacity constraints. The coherence of the stages is ensured by the fact that the volume of raw materials sent for processing is determined by the results of the first stage and serves as the resource base for the second stage. It is also necessary to consider possible production options related to external factors and production-economic conditions.

The modeling algorithm includes sequential procedures: defining the goal (usually maximizing profit or income while taking costs into account), creating a system of constraints, and preparing the initial data. To account for uncertainty, key parameters are defined at condition levels, allowing for alternative optimal plans and assessing the range of possible outcomes. First, the production model is implemented for each scenario, and the resulting raw material volumes are transferred to the processing model. The next step is to verify the consistency of the decisions. If there is a shortage of raw materials or resources (unfavorable conditions of the first stage), corrective actions are taken (purchasing raw materials, redistributing flows between sales and processing, changing product types and output volumes). Under favorable conditions, the company can obtain additional products for sale and processing.

The result is the creation of coordinated production and processing plans for each condition level, specifying the expected economic outcome and key constraints. The proposed algorithm for implementing a multi-stage, multi-level parametric model is based on data from the agricultural Joint-Stock Company «Belorechenskoye». Optimal production and processing plans were obtained for unfavorable, average, and favorable conditions, and constraints that most significantly impact the outcome and require management adjustments were identified.

TIME SERIES CLUSTERING FOR THE ANALYSIS OF BIOMASS PYROLYSIS TG DATA

Bachurin Roman, Zakharova Irina

University of Tyumen
Tyumen, Russia
romanbachurin@gmail.com

The study of the thermal behavior of biomass during pyrolysis is a key task in the development of technologies for processing renewable feedstocks into valuable products. Thermogravimetric analysis (TG) serves as the primary experimental method for recording mass loss during sample heating [1]. However, the accumulation of large experimental datasets, represented by multipack curves with varying decomposition step positions, creates significant difficulties for visual analysis and interpretation.

This paper presents an automated approach to the analysis of TG curves from pyrolysis of plant and animal biomass based on time series clustering. The study was performed on a dataset of 164 experimental curves obtained by heating samples from 26 to 1000 °C at a rate of 10 °C/min in an inert atmosphere. Each time series contained an average of 4850 data points.

The main methodological challenge was the selection of an appropriate similarity measure for curves characterized by nonlinear distortions along the temperature axis due to differences in the kinetics of component decomposition. Several metrics were compared: Euclidean distance, dynamic time warping (DTW), Soft-DTW, derivative DTW, cross-correlation, cosine similarity, and matrix profile. The DTW metric [2], which provides optimal nonlinear alignment of sequences, demonstrated the best performance, achieving a maximum silhouette coefficient of 0.534 [3].

For clustering, a modified k-means [4] algorithm with DTW centroids (DTW Barycentric Averaging) was applied. The optimal number of clusters k=4 was determined based on a combined analysis of the elbow method and silhouette coefficient.

The resulting clustering enabled the identification of seven statistically significant groups of curves reflecting distinct types of thermal behavior of the investigated samples (Fig. 1). The obtained sample distribution correlates with the known physicochemical properties of the studied materials, confirming the validity of the proposed method.

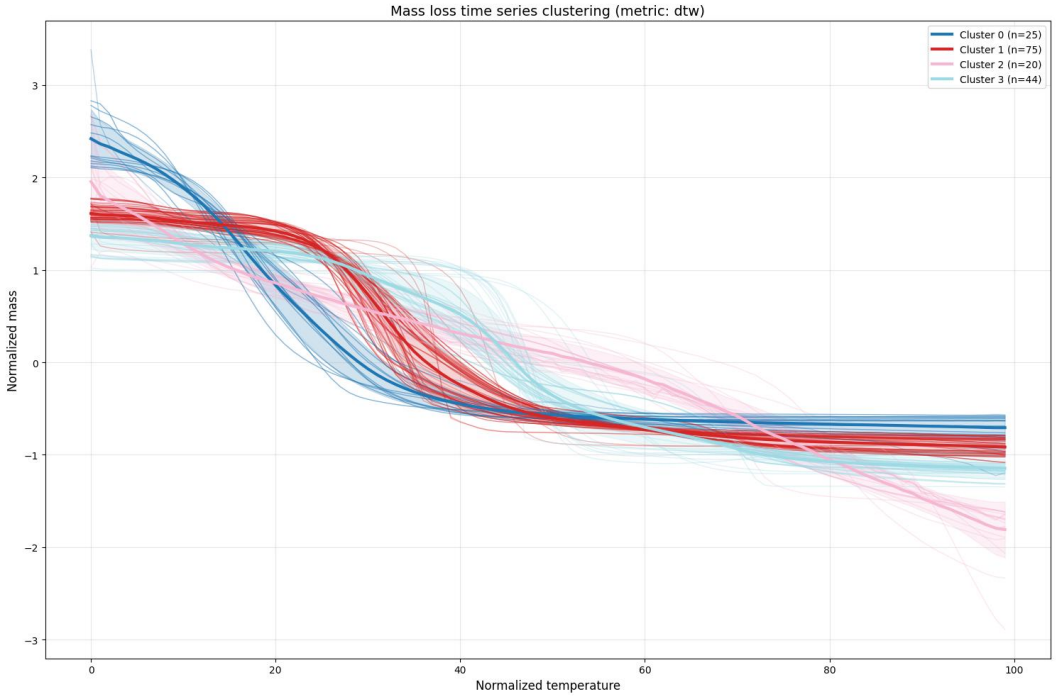


Fig. 1. The plot of normalized TG curves and their cluster centroids

The developed approach automates the classification of experimental data, enables the identification of reference curves for modeling (based on cluster centroids), and facilitates the detection of anomalies associated with changes in feedstock composition or measurement errors. Further development of the approach involves the use of the rate of mass loss as an additional metric and the application of physics-informed neural networks (PINNs) for the interpretation of cluster structures.

Acknowledgements. This study was supported by the Ministry of Science and Higher Education of the Russian Federation, project FEWZ-2024–0052.

References

1. El-Sayed, S.A., Khass, T.M. & Mostafa, M.E. Thermal degradation behaviour and chemical kinetic characteristics of biomass pyrolysis using TG/DTG/DTA techniques. Biomass Conv. Bioref. 14, 17779–17803, 2024, doi: 10.1007/s13399-023-03926-2.
2. Shuyang Zhang, Lun Zhang, Xiaosong Zhang, Clustering based on dynamic time warping to extract typical daily patterns from long-term operation data of a ground source heat pump system, Energy, Volume 249, 123767, ISSN 0360-5442, 2022, doi: 10.1016/j.energy.2022.123767.
3. I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, “Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta”, JAIC, vol. 7, no. 1, pp. 95–103, 2023, doi.org/10.30871/jaic.v7i1.4947.
4. Vera, J.F., Angulo, J.M. An MDS-based unifying approach to time series K-means clustering: application in the dynamic time warping framework. Stoch Environ Res Risk Assess 37, 4555–4566, 2023, doi: 10.1007/s00477-023-02470-9.

DIGITAL TECHNOLOGY OF AUTOMATION FOR AGENT MODELING OF MICROGRID INTERACTIONS

Chekan Mikhail

Matrosov Institute for System Dynamics and Control Theory of SB RAS
Irkutsk, Russia
chekoopa@mail.ru

Nowadays, the study of the operation and development of local energy systems on the Baikal Natural Territory is a challenge [1]. An important class of these systems are microgrids. In this context, modelling a possible interaction of microgrids in terms of the energy resource exchange to increase their resilience is particularly important.

Unfortunately, the known methods and tools for conducting such studies [2] do not support the development, modification, and adjustment of economic mechanisms to manage the supply and demand for energy resources. Furthermore, the social aspect of microgrid interactions under conditions of large external disturbances has not been taken into account. The study of the functioning and development of energy systems often involves solving the problem of structural and parametric optimization [1], which is characterized by the need to analyze a large number of energy resources production, transportation, and consumption objects, their relations, and various probabilistic distributions of random variables. This determines the high complexity of the problem to be solved.

Therefore, the use of digital twins [3, 4] and multi-agent technologies [5] are promising approaches to solving this problem. In particular, digital twins allow us to analyze large sets of current and historical data on the state of objects and implement feedback and control mechanisms. At the same time, multi-agent technologies provide improved control through the self-organization of agents that combine their cooperation to solve common problems with the competition to achieve individual goals.

In the report, a digital technology for creating multi-agent systems to model a coalition of interconnected microgrids is represented. The distinguishing feature of this technology is its ability to support the development, modification, and adjustment of economic mechanisms for managing supply and demand for electrical resources, taking into account social aspects of microgrid interaction under conditions of large external disturbances.

A toolkit has been developed for implementing this technology [6]. It represents a special add-on the Java Agent DEvelopment Framework (JADE) [7]. The general modeling scheme is shown in Fig. 1. The crucial part is JADE environment with multi-agent system representing a microgrid coalition with two-level energy trading based on a virtual coordination agent. A domain expert interacts with it using the add-on, which provides configuration tools, e.g. economical mechanism configurator and visual agent behavior constructor based on hierarchical state machines.

The technology was tested on model cases with forecasting of external disturbances using meteorological data. Experimental results have shown a significant reduction in overheads for developing agents and economic mechanisms using the toolkit. The reduction in overhead ranged from 20% to 45% compared to creating a multi-agent system using JADE without the toolkit. It is important to note that these advantages are more clearly manifested as the number of agents grows and the frequency of modifications to economic mechanisms increases. The reliability of the modeling results obtained is confirmed by comparing them with calculations in MATLAB.

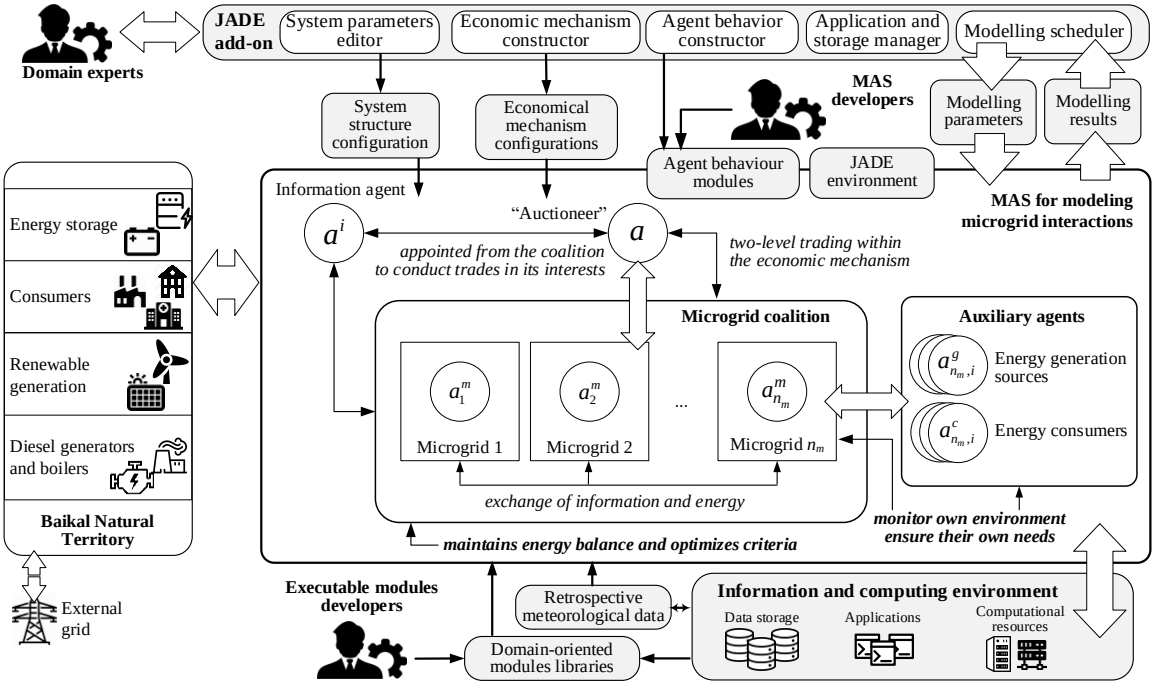


Fig. 1. General modeling scheme

Acknowledgements. The study was supported by the Ministry of Science and Higher Education of the Russian Federation, project no. FWEW-2026-0012.

References

1. Edeleva O., Edelev A., Voskoboinikov M., Feoktistov A. Scientific Workflow-Based Synthesis of Optimal Microgrid Configurations // Energies. 2024, Vol. 17, no. 23. P. 6138. DOI: 10.3390/en17236138.
2. Ahmadi S., Saboohi Y., Vakili A. Frameworks, quantitative indicators, characters, and modeling approaches to analysis of energy system resilience: A review // Renewable and Sustainable Energy Reviews. 2021. Vol. 144. – P. 110988. DOI: 10.1016/j.rser.2021.110988.
3. Massel L., Massel A. Smart digital twins as a trend of energy systems intellectualization in Russia // Lecture Notes in Networks and Systems. 2023. Vol 777. P. 334–346. DOI: 10.1007/978-3-031-43792-2_32 2023.
4. Kosyakov D.V., Marchenko M.A. Digital Twins in Large-Scale Scientific Infrastructure Projects // Supercomputing Frontiers and Innovations. 2023. Vol. 10, no. 3. P. 88–106. DOI: 10.14529/jsfi230308.
5. Stennikov V., Barakhtenko E., Mayorov G., Sokolov D., Zhou B. Coordinated management of centralized and distributed generation in an integrated energy system using a multi-agent approach // Applied energy. 2022. Vol. 309. P. 118487. DOI: 10.1016/j.apenergy.2021.118487.
6. Bychkov I.V., Feoktistov A.G., Chekan M.A. Modeling Agent Behavior in Interacting Microgrids // Proceedings of the 10th International Conference on Information Technology and Nanotechnology (ITNT 2024). IEEE, 2024. P. 1–5. DOI: 10.1109/ITNT60778.2024.10582380.
7. Bellifemine F., Poggi A., Rimassa G. JADE: A FIPA2000 compliant agent development environment // Proc. of the 5th Intern. Conf. on Autonomous Agents (AGENTS'01). New York: ACM, May 2001. P. 216–217. DOI: doi.org/10.1145/375735.376120.

MODELING THE POPULATION OF LIVESTOCK IN RELATION TO THE NATURAL-ECONOMIC ZONES OF MONGOLIA

Enkhbat Ariunjargal, Ivanyo Yaroslav

Irkutsk State Agrarian University named after A.A. Ezhevsky
Molodezhny, Russia
eariuka329@gmail.com

Mongolia's territory is divided into three natural-economic zones and six main natural-economic regions: Western, Khangai, Eastern, Central, Gobi, and Ulaanbaatar. Most researchers combine the Gobi and Central regions into the Central Region. Although Ulaanbaatar is considered a separate region, its quantitative indicators have little connection to agriculture. Therefore, it is usually not considered in isolation.

Depending on natural conditions, certain household specializations have traditionally developed within regions based on livestock species. In the Western region, due to its arid climate and vast pastures, sheep, goats, camels, and horses are primarily bred. The Khangai region is favorable for the development of cattle, sheep, and horse breeding. Camel breeding, small-scale livestock farming (sheep and goats), and cattle are concentrated in the Central region. Horse breeding, beef cattle breeding, and sheep breeding are well developed in the Eastern region. Furthermore, there is high potential for feed production.

One of the serious problems of pasture livestock farming is pasture degradation. Its severity varies across different regions. Pasture degradation is most severe in the Central region.

To optimize livestock numbers, a linear model of specialization based on pasture conditions was developed. The model's objective function is aimed at maximizing the farm's overall profit, taking into account the specific natural-economic conditions. The model considers the profit from the sale of various animal species, their productivity, feed, water, and labor requirements, as well as pasture area limitations and land degradation. Additionally, limits on the number of animals for individual species were introduced. This allows for the modeling of various farm strategies, the structure of different animal species, and their numbers.

The proposed model, taking into account the specifics of production in different natural and economic regions, allows for the development of various options for optimizing livestock structure and numbers. In particular, the degradation coefficient, introduced in one of the constraints, varies widely. For some regions, it is close to zero, while for others, it can reach 0.33.

On the one hand, there is pasture degradation, while on the other, there is an increase in forage crop yields. In terms of yield volume and growth, the Khangai and Central regions have the greatest potential. The Eastern and Western natural and economic regions are somewhat less successful in this regard.

The high regression relationship between yields over time facilitates the use of parametric programming for planning farm specialization based on the structure and number of livestock. Modeling the dynamics of forage crop yields for different natural and climatic regions at different levels of the time series hierarchy reveals the potential of parametric programming for favorable, unfavorable, and average situations.

Thus, by assessing pasture livestock farming traditions in different regions of Mongolia, the natural and climatic conditions that determine the availability of feed and water for animals, and the level of pasture degradation, adequate variants of a model for optimizing the structure and number of livestock were developed to improve management decisions.

ARCHITECTURE OF A MULTI-AGENT DISTRIBUTED CONTROL SYSTEM FOR THE FUNCTIONING OF AN ACCOUNTING INFORMATION SYSTEM

Erzhenin Roman

Krista-Irkutsk LTD
Irkutsk, Russia
rerzhenin@gmail.com

The use of an agent-based approach in the development of accounting information system (AIS) management support systems is one of the most relevant areas at the intersection of control theory, artificial intelligence, and financial management theory.

Figure 1 shows an improved AIS management model, which utilizes user behavior data as a tool for improving the efficiency of user-software interaction.

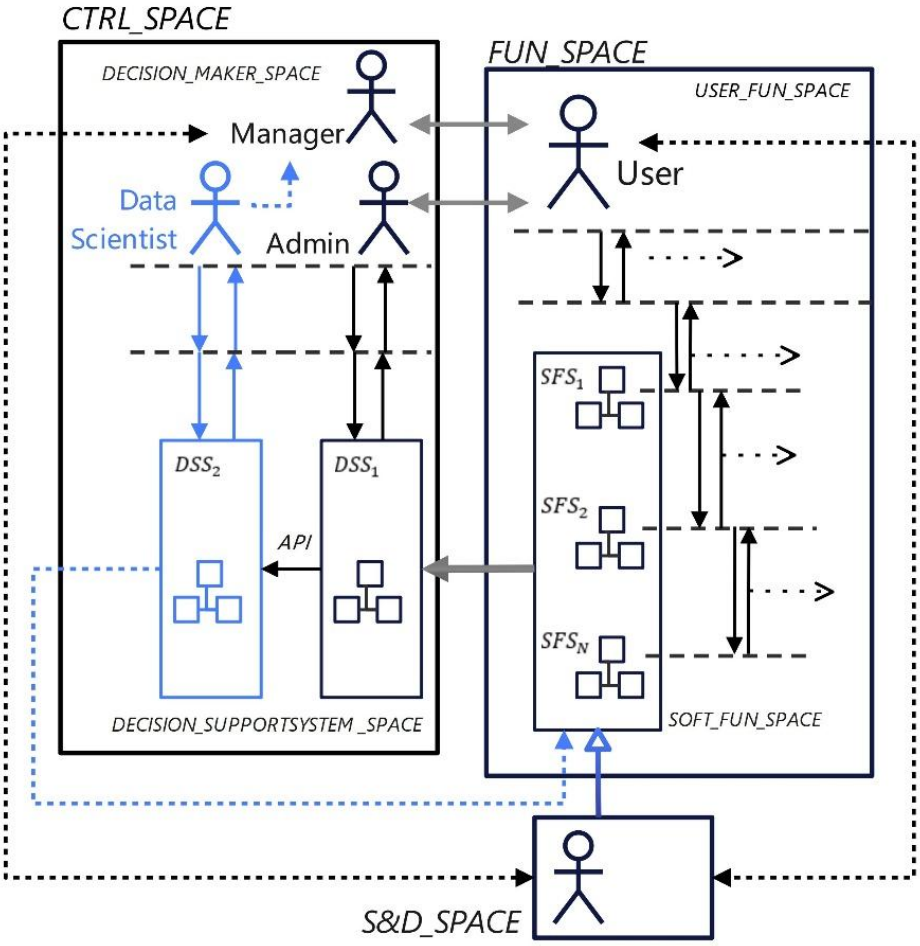


Fig. 1. Model for managing the functioning of the penal system

In the schematic model presented, the control space E is conditionally divided into three functional subspaces, each of which is characterized by its own groups of actors with their own connections and actions

$$E = \{CS, FS, SDS\}$$

Where *CS (CTRL_SPACE)* is the IS operation and development management space; *FS* (*FUN_SPACE*) is the IS functioning space; and *SDS* (*S&D_SPACE*) is the support and development space.

In turn, the IS functioning space *FUN_SPACE* is divided into two subspaces.

$$FS = \{UFS, SFS\}$$

where *UFS* (*USER_FUN_SPACE*) is the user fun space; *FS* (*SOFT_FUN_SPACE*) is the application fun space.

The *USER_FUN_SPACE* user fun space includes specialists performing duties in their functional areas

$$UFS = \{UFS_1, UFS_2 \dots UFS_N\}$$

The software functioning space *SOFT_FUN_SPACE* includes various groups of software components that interact with specialists to achieve their goals and solve functional problems

$$SFS = \{SFS_1, SFS_2 \dots SFS_N\}$$

The control space for the functioning and development of the *CTRL_SPACE* information system is also divided into two subspaces

$$CS = \{DMS, DSS\}$$

where *DMS* (*DECISION_MAKER_SPACE*) is the decision-maker's space; *DSS* (*DECISION_SUPPORTSYSTEM_SPACE*) is the decision support system's operational space.

The *DECISION_MAKER_SPACE* includes a group of decision-makers in accordance with their assigned functions – administrators, managers at various levels of resource management, and managers of the information system's operational results

$$DMS = \{DMS_1, DMS_2 \dots DMS_N\}$$

The operational space of the decision support system *DECISION_SUPPORTSYSTEM _SPACE* includes various components and software systems on the basis of which decisions are made on IS management

$$DSS = \{DSS_1, DSS_2 \dots DSS_N\}$$

The support and development space of the functional software system *S&D_SPACE* includes such stakeholder groups as “technical support” (*SP*), “development” (*DV*) and “management” (*MN*)

$$SDS = \{SP, DV, MN\}$$

In the current architecture of a modern accounting system, user behavior data is used very sparingly, primarily to resolve incidents involving disputes resulting from user errors or algorithmic errors. At the same time, behavioral analytics is a source of knowledge that, with a specialized methodology, can be an effective tool for influencing both unskilled and skilled users, facilitating understanding of the subject matter and the dynamic evolution of accounting requirements and rules within the accounting system

Using the methods of constructing an abstract logical model, we will represent the subspace DSS_2 as a multi-agent distributed system (MADS)

$$E = \{O, S, \Omega\}$$

where O – agent environment objects; S – connections between environmental objects O; Ω – actions on objects in subspace DSS_2.

MADS objects may include the following:

- Observer agents, which monitor event logs and logs;
- Analyst agents, which analyze data for deviations from typical behavior;
- Performer agents, which perform actions based on the analysis results (hints, recommendations, consultations);
- Coordination agents, which ensure synchronization between agents;
- Arbitrator agents, such as a Data Scientist, which makes decisions in complex cases using business logic rules or machine learning;
- Exchange infrastructure for asynchronous interaction between agents;
- Knowledge base, which stores behavior patterns, behavioral analytics rules, response templates for behavioral deviations, domain knowledge, etc.

The development of MADS will enhance the intelligence of the management space for the functioning and development of the information system, reduce lost time, and improve the efficiency of the software system.

AUTOMATION OF REQUEST PROCESSING BASED ON AI AND STOCHASTIC MODELS

Fedotov Nikita

Baikal State University
Irkutsk, Russia
010686@bgu.ru

The relevance of this work is driven by the need to improve the efficiency of processing large volumes of requests in modern business processes. Companies across various industries face a growing flow of customer inquiries on a daily basis, which requires the optimization of processing procedures to ensure competitiveness and service quality.

The aim of the work is to develop methods for automating request processing based on the integration of artificial intelligence technologies and mathematical models of queueing theory.

To achieve this goal, the following tasks were completed: existing approaches to request processing automation were analyzed; a hybrid system architecture was developed; a mathematical model of the request processing procedure was constructed; the effectiveness of the proposed solutions was evaluated.

The scientific novelty of the study lies in the development of a comprehensive approach to request processing automation that combines intelligent classification with resource optimization based on queueing theory.

The proposed system architecture includes three main components: an intelligent processing module for automatic classification and prioritization of requests based on the transformer architecture [3], a queue management module for routing requests among operators, and an analytical module for forecasting and optimizing processes based on queueing models.

For performance analysis, a mathematical model was constructed based on a multi-server queueing system with a Markovian arrival process and phase-type service time distribution (MAP/PH/c) with priorities [1, 2]. The model allows calculating the key system characteristics: average waiting time, queue length, and service denial probability.

The application of the transformer architecture for processing textual data with minimal fine-tuning allowed the system to adapt to various request types without the need for extensive retraining.

References

1. Kleinrock L. Queueing Systems. Volume 1: Theory. New York, Wiley, 1975. 417 p.
2. Neuts M. F. Structured Stochastic Matrices of M/G/1 Type and Their Applications. New York, Marcel Dekker, 1989. 512 p.
3. Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need. Advances in Neural Information Processing Systems, 2017, vol. 30, pp. 5998–6008.

ARTIFICIAL INTELLIGENCE TECHNOLOGIES TO INCREASE THE ENERGY EFFICIENCY OF ECONOMIC SECTORS

Galperova Elena, Galperov Vasilii

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
galper@isem.irk.ru

Increasing energy efficiency helps solve three critical challenges in modern economic development: (1) reducing energy consumption, (2) improving energy security, and (3) reducing the negative impact on the environment.

The use of artificial intelligence (AI) technologies is currently becoming a global trend in energy management, reducing all types of losses, and improving the economic and environmental performance of enterprises and industries [1].

Analysis of the main areas of AI technology application in individual economic sectors to improve their energy efficiency is becoming an important step in developing long-term energy consumption forecasts.

In industry, the implementation of AI can improve energy efficiency by optimizing the composition and operating time of equipment, rationalizing energy costs, and accelerating the development and launch of new products [2].

In transportation, AI enables energy savings through real-time traffic management, route optimization, and predicting maintenance schedules [3].

In buildings, AI technologies enable energy management, loss detection, optimization of equipment operation and repair, and efficient interaction with the power grid [3].

Long-term forecasting of the dynamics of the volume and structure of energy consumption has been and remains the first and most important stage in the development of strategies and programs for the development of the country's energy sector and economy.

Acknowledgements. The research is carried out as part of the State Assignment Project No. FWEU-2026-0009 of the Basic Research Program of the Russian Federation for 2021-2030.

References

1. The future of AI and energy efficiency. https://www.ibm.com/think/insights/future-ai-energy-efficiency
2. Optimize Efficiency With AI-Driven Energy Management. https://www.pecan.ai/blog/optimize-efficiency-with-ai-energy-management/
3. Energy and AI. World Energy Outlook Special Report. International Energy Agency. 2025. 304 p.

GEOMAGNETIC INDUCED CURRENTS AND IONOSPHERIC DISTURBANCES AS RISK FACTORS FOR ENERGY SYSTEMS

Garashchenko Aleksandr[1], Tomin Nikita[2], Dreglea Alena[1], Sidorov Denis[2]

[1]National Research Technical University
[2]Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
garashchenkoaa@mail.ru

The aim of this work was to comprehensively analyze the relationship between GICs and ionospheric disturbances, and to develop approaches for their forecasting based on machine learning and statistical methods.

A total of 800 GIC scalograms were used (400 each for the "GIC" and "geomagnetically quiet intervals" classes) to train seven classifiers, including Logistic Regression and Gaussian Naive Bayes. Ionospheric parameters—TEC variations, the ROTI index, and calibrated TEC—were obtained for 21 intervals using data from the SIMuRG system.

GIC classification based on the scalogram data was performed (Table 1). Logistic Regression achieved the highest overall classification accuracy (97.5%). The Gaussian Naive Bayes classifier demonstrated no false positive detections (FP=0); however, its recall was only 0.87 (missing 9 GIC events).

The relationship between GICs and ionospheric parameters derived from GNSS data was assessed. Statistical analysis revealed no linear correlation between TEC variations and GICs. A very weak, yet statistically significant, correlation was found for ROTI and calibrated TEC (coefficients < 0.25), indicating a nonlinear and indirect relationship. A threshold effect was observed for ROTI: GIC intensity is maximal when ROTI < 20 and decreases when ROTI ≥ 20.

Table 1. Best accuracy performance of GIC image classifiers

Classifier	True Positives	False Positives	False Negatives	True Negatives	Overall Accuracy
Logistic Regression	69	4	0	87	0.975
Gaussian Naive Bayes	60	0	9	91	0.944

Acknowledgements. The work was carried out within the framework of the state assignments FWEU-2026-0012, FWEU-2026-0015. The authors express their gratitude to Aksenovich T. and Selivanov V. for providing the scalogram images.

References

1. Aksenovich T., Selivanov V. A Deep Learning Approach for the Automated Classification of Geomagnetically Induced Current Scalograms // Applied Sciences. 2024. V. 14. N 2. P. 895.
2. Yasyukevich Y.V., Kiselev A.V., Zhivetiev I.V., Edemskiy I.K., Syrovatskii S.V., Maletckii B.M., Vesnin A.M. SIMuRG: System for ionosphere monitoring and research from GNSS // GPS solutions. 2020. V. 24. P. 1-12.

ON DEFECTS OF WELDED JOINTS IN STEEL PENSTOCKS OF HYDROELECTRIC POWER PLANTS

Gilmutdinov Vladimir[1,2], Popov Maksim[2]

[1]Melentiev Energy Systems Institute SB RAS
[2]Irkutsk National Research Technical University
Irkutsk, Russia
gilmutdinovvr@ex.istu.edu

Steel welded metal structures play a key role in the construction and energy sectors, especially in hydraulic engineering structures such as hydroelectric power plant penstocks. Statistics show that a significant portion of penstock failures are related to the quality of welded joints. These structures operate under high pressure, cyclic loads, vibration, and corrosive environments, which requires a systematic approach to monitoring their condition.

The aim of this work is to systematize knowledge about weld defects in the context of HPP penstock operating conditions and to analyze their impact on the durability of the structures.

Weld defects are classified by location (surface and internal) and geometry (planar defects – cracks and lack of fusion; volumetric defects – pores and slag inclusions). Hot and cold cracks, as well as defects in the heat-affected zone where metal softening occurs, pose a particular danger.

The most critical defects for HPP penstocks are lack of fusion and cracks in the weld root, as this zone experiences the maximum tensile stresses. Undercuts, as stress concentrators, significantly reduce cyclic durability. Cold and corrosion cracks are characteristic of thick-walled steels in contact with water. Chains of pores and slag inclusions, combined with vibration, reduce the effective cross-section of the weld and provoke fatigue phenomena.

Non-destructive testing methods applied to HPP penstocks are specific. Inspection requires the use of various methods, for example, ultrasonic testing, under conditions of limited access. Visual and measurement inspection, such as flaw detection, are used to identify surface defects. During operation, periodic testing and thickness measurement are particularly valuable for assessing wear.

Conclusion: the main reason for the reduction in the service life of HPP penstocks is the presence of welding defects acting as stress concentrators. Lack of fusion in the weld root, undercuts, and cracks pose the greatest danger. Priority diagnostic measures should be focused on their detection and monitoring.

Acknowledgements. The work was carried out within the framework of the state task, project FZZS-2024-0003. The authors are grateful to Roman Vladimirovich Kononenko for setting the problem.

FACTOR ANALYSIS OF RESPIRATORY TRACT INJURY RISKS USING THE SHEPARD OPERATOR

Gornov Aleksandr, Anikin Anton, Lykhin Semen

Matrosov Institute for System Dynamics and Control Theory of SB RAS
Irkutsk, Russia
semen@icc.ru

Modern data analysis relies on the use of effective approximation and interpolation methods, taking into account high dimensions and the absence of a regular grid. One such approach is the Shepard function data analysis technology, a universal interpolation method based on the weighted average of known points, taking into account their distance from the interpolation point. The work is devoted to the experimental study of the properties of the Shepard approximant and its modifications (with the addition of a derivative and smoothing components). To implement the proposed methodology in the Python programming language, it is the classical form of the Shepard operator ([1, 2]) – a fragment of the program code is shown in the table 1.

Table 1. Software implementation of the classic version of the Shepard operator in Python

```
def shepard_interpolation(x, x_points, y_points, p=4):
    distances = np.linalg.norm(x_points - x, axis=1)
if np.any(distances == 0):
        return y_points[np.argmin(distances)]
    weights = 1 / (distances ** p + 1e-10)
    denominator = np.sum(weights)
if denominator == 0:
        return np.nan
    numerator = np.sum(weights * y_points)
    return numerator / denominator
```

The results of application of the program realization of the Shepard operator for minimizing two-dimensional nonconvex function $f(x,y) = \sin\big(\pi x(1+0.3y)\big) + 0.15xy + 0.1(x+2y)^2$ is presented on Fig. 1.

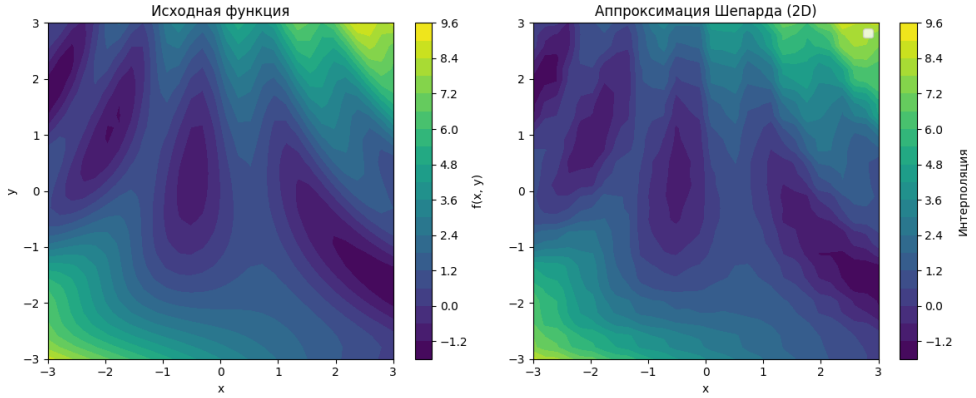


Fig. 1. The numerical results of the Shepard's approximation two-dimensional nonconvex function

The implemented data approximation technology using the Shepard operator allowed for factor analysis of data in the task of assessing the risks of developing chronic diseases of the upper respiratory tract.

References

1. Shepard D. A Two-Dimensional Interpolation Function for Irregularly-Spaced Data // Proceedings of the 1968 ACM National Conference, стр. 517–524.
2. William J. Gordon and James A. Wixom. Shepard's Method of "Metric Interpolation" to Bivariate and Multivariate Interpolation. Mathematics of computation, 32(141) 1978, P. 253–264

THE MAIN DIRECTIONS OF QUALITY CONTROL OF PROJECT DOCUMENTATION IN AVIATION INDUSTRY

Govorkov Aleksei

Irkutsk National Research Technical University
Irkutsk, Russia
govorkov_as@ex.istu.edu

Abstract. The research focuses on developing a methodology for quality control of project documentation in the aviation industry, taking into account human resources' qualitative indicators. The study addresses the critical need for high-quality design documentation in aircraft manufacturing.

Introduction. The aviation industry's design process requires highly qualified personnel with extensive experience. The quality of project documentation directly impacts aviation equipment production. Timely quality control and appropriate staff selection are crucial for successful project implementation.

Research Methodology. The study employs a comprehensive approach to quality control, incorporating:

- systematic assessment of personnel qualifications;
- multi-stage monitoring of design processes;
- objective evaluation methods;
- automated selection system for personnel.

The methodology evaluates personnel based on seven key criteria:

- educational background;
- professional experience;
- documentation development speed;
- error rate in previous projects;
- health records (sick leaves);
- vacation schedules;
- current workload.

Main Findings. The developed quality control system provides:

- objective personnel evaluation;
- optimal team formation;
- efficient workload distribution;
- improved documentation quality;
- 15-17% time reduction;
- 28-40% cost optimization.

Discussion. The research introduces an innovative personnel assessment system using a comprehensive scoring approach. The methodology considers both professional and personal employee characteristics. The application of the Laplace criterion ensures unbiased decision-making in personnel selection.

Practical Implications. The proposed system significantly enhances project documentation development processes through:

- standardized quality control procedures;
- objective personnel assessment;
- efficient resource allocation;
- risk minimization in project implementation.

Conclusion. The developed methodology demonstrates high effectiveness in improving project documentation quality. Its implementation leads to:

- enhanced production efficiency;
- reduced error rates;
- optimized resource utilization;
- improved project outcomes.

The methodology's adaptability allows its application in various industries requiring high-quality design documentation.

References

1. Chernousenko R.E. Personnel management system at an industrial enterprise. International Journal of Humanities and Natural Sciences, 2019, vol. 12-3 (39), pp. 194-196.
2. Tregulova N.G. Features of strategic personnel management in an industrial enterprise. Regional problems of economic transformation, 2018, no. 4, pp. 204-209.
3. Romanova M.M. Features of personnel management in the project organization of activities. Bulletin of Omsk University, 2011, no. 2, pp. 30-34.
4. Bandurin A.V. Features of a project-oriented company as an environment for the manifestation of personnel competencies. Creative economy, 2023, vol. 17, no. 5, pp. 1671-1688.
5. Chulanova O.L. et al. Project competencies of personnel in the conditions of the BANI world. Bulletin of Eurasian Science, 2022, vol. 14, no. 4.
6. Sultanova F.R. Role of motivational and value orientations of employees in assessing the attractiveness of organizational culture. Dissertation, Moscow, 2023, 163 p.
7. Leonova A.B., Sultanova F.R. Motivational and value orientations of personnel and attractiveness of organizational culture. Questions of psychology, 2018, no. 4, pp. 80-92.
8. Kostina E.V. Evaluation of conflict behavior strategies and communicative tolerance. Bulletin of Kostroma State University, 2022, vol. 28, no. 3, pp. 100-107.
9. Demin G.A. Methods of managerial decision-making. Perm State University, 2019, 88 p.
10. Yares O.B., Panshin I.V. Methods of managerial decision-making. Vladimir State University, 2011, 66 p.

A HIERARCHICAL GRAPH-BASED MODEL OF DOCUMENT STRUCTURE FOR CONTROLLED LECTURE SUMMARY GENERATION

Ismagulov Milan

Yugra State University
Khanty-Mansiysk, Russia
m_ismagulov@ugrasu.ru

Introduction. The task of generating structured lecture summaries from educational content has become increasingly relevant with the widespread adoption of large language models (LLMs). While modern LLMs demonstrate strong capabilities in text generation, they often lack explicit control over document structure, leading to unstable hierarchy, loss of logical coherence, and excessive fragmentation in long-form outputs [3].

Most existing approaches rely on either universal document templates or implicit structural inference performed by the language model itself. These strategies do not adequately reflect real hierarchical patterns inherent in documents from different subject domains. As a result, generated summaries often diverge from domain-specific structural conventions.

This work proposes a hierarchical document structure model based on Hierarchical Document Structure Analysis (HDSA) and graph theory, enabling explicit extraction, evaluation, and reuse of structural patterns in the form of formalized JSON templates for controlled text generation [2].

Hierarchical Document Structure Analysis Model. A document is represented as a rooted hierarchical graph, where nodes correspond to structural elements such as headings, paragraphs, lists, formulas, and figures, and edges encode parent-child relations reflecting semantic nesting (Fig. 1).

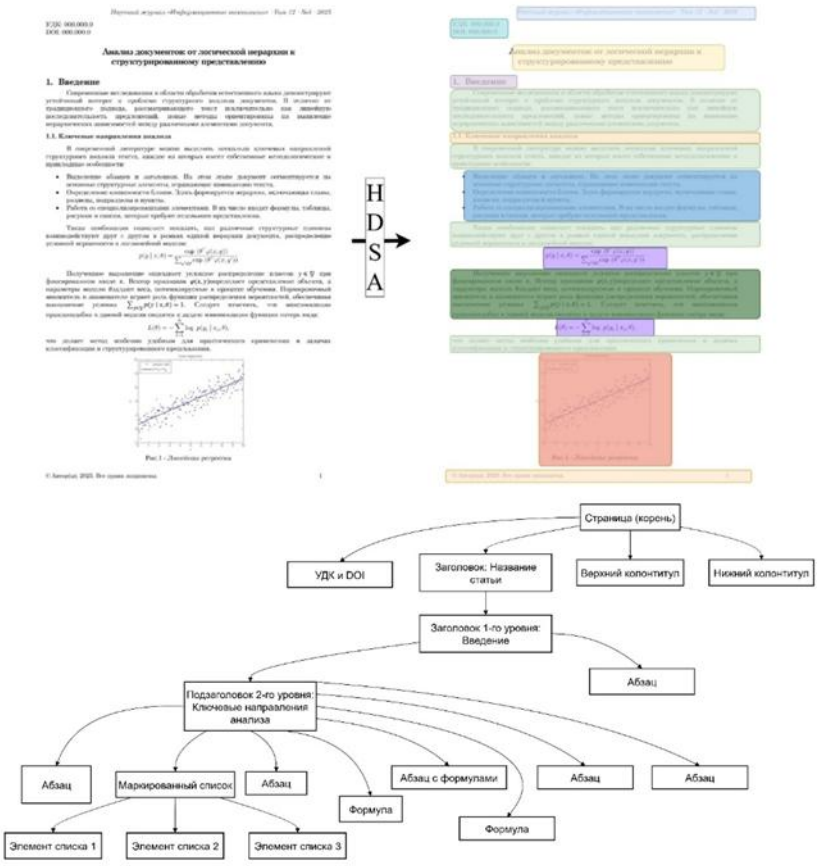


Fig. 1. Example of document segmentation into structural blocks using the HDSA method and the corresponding hierarchical tree representation

Unlike classical Document Layout Analysis (DLA), which focuses on spatial arrangement and visual reconstruction, HDSA operates at the logical-semantic level, modeling how content blocks are

organized conceptually rather than geometrically. This distinction makes HDSA particularly suitable for downstream text generation tasks [1].

Each node in the document graph is characterized by:

1. Structural type (e.g., heading, paragraph, equation).
2. Depth relative to the document root.
3. Connectivity within the hierarchy.

This representation allows the document to be analyzed as a formal graph object, enabling the application of quantitative structural metrics.

Graph-Based Structural Metrics. To capture stable structural characteristics of real documents, several graph-theoretic measures [5] are introduced:

- Average hierarchical depth, reflecting the effective nesting complexity of the document.
- Content diversity measure, quantifying the balance between different block types (text, formulas, lists, figures).
- Text-centricity ratio, estimating the proportion of textual content within the overall structure.
- Betweenness centrality, used to identify structurally critical elements such as section headings that connect multiple semantic branches.

These metrics provide a compact yet expressive description of document structure and serve as a basis for comparing documents across subject domains. Template Construction via Structural Analysis. Based on extracted structural metrics, the study introduces a method for synthesizing JSON-based structural templates [4]. Each template encodes:

- reference metric values with semantic interpretation,
- allowed heading hierarchy levels and their relative importance,
- admissible content block types and formatting constraints,
- stylistic requirements for generated text.

Importantly, the core hierarchical schema remains invariant across domains, while differences between subject areas are reflected in content diversity and block composition rather than in heading depth or tree topology. This design ensures both structural stability and domain adaptability during generation.

Experimental Validation (Brief Overview). The proposed model was evaluated on lecture materials from three subject domains: mathematics, law, and ecology. For each domain, document corpora were analyzed to extract structural metrics and derive corresponding templates.

Controlled text generation experiments were conducted both with and without templates. The results demonstrate that template-guided generation significantly reduces structural variability and eliminates degenerate cases such as loss of modality-specific elements (e.g., mathematical expressions). At the same time, the experiments reveal a systematic tendency of LLMs toward excessive fragmentation in structurally rich domains, highlighting intrinsic limitations of current generative mechanisms rather than shortcomings of the proposed model [3].

Discussion. The findings confirm that explicit structural modeling provides tangible benefits for long-form text generation. By separating structural control from semantic generation, the proposed approach improves reproducibility and interpretability of generated documents.

Crucially, the method aligns well with multi-agent and multi-model processing pipelines, where structural templates can act as coordination artifacts between specialized agents responsible for different modalities or processing stages.

The observed over-fragmentation effect suggests that future work should focus on adaptive aggregation rules and hierarchical regularization strategies within template definitions.

Conclusion. This work presents a graph-based hierarchical document structure model grounded in HDSA and graph theory. The model enables extraction of stable structural patterns from real documents and their transformation into formal JSON templates for controlled text generation.

Experimental results across multiple subject domains demonstrate that template-guided generation stabilizes document structure without sacrificing content completeness. The approach provides a practical and interpretable alternative to monolithic multimodal generation and supports scalable integration into multi-agent systems for lecture summarization [1, 5].

Acknowledgements. The author expresses gratitude to Yugra State University for providing the academic and research environment in which this study was conducted, and for the institutional support during the author's study and professional activity.

References

1. Wang J., Hu K., Zhong Z., Sun L., Huo Q. Detect-order-construct: A tree construction-based approach for hierarchical document structure analysis // Pattern Recognition. 2024. Vol. 156. Art. no. 110836. DOI: 10.1016/j.patcog.2024.110836.
2. Giovannini S., Marinai S. A survey on reading order, table of contents, and structure extraction in document analysis // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2025. P. 7585–7594.
3. Liu Y., Li D., Wang K., Xiong Z., Shi F., Wang J., Li B., Hang B. Are LLMs good at structured outputs? A benchmark for evaluating structured output capabilities in LLMs // Information Processing & Management. 2024. Vol. 61, no. 5. Art. no. 103809. DOI: 10.1016/j.ipm.2024.103809.
4. Lu Y., Li H., Cong X., Zhang Z., Wu Y., Lin Y., Liu Z., Liu F., Sun M. Learning to generate structured output with schema reinforcement learning // Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics. 2025. P. 4905–4918.
5. Chen Y., Liu R., Zheng J., Wen D., Peng K., Zhang J., Stiefelhagen R. Graph-based document structure analysis [Electronic resource]. 2025. Available at: https://arxiv.org/abs/2502.02501 (accessed: 03.02.2025).

APPLICATIONS OF MULTILEVEL MODELING IN AGRICULTURAL PRODUCTION MANAGEMENT

Ivanyo Yaroslav, Petrova Sofia

Irkutsk State Agrarian University named after A.A. Ezhevsky
Molodezhny, Russia
sofia.1987@bk.ru

In this paper, multilevel modeling refers to the process of obtaining enterprise performance results under average, unfavorable, and favorable conditions, as well as during extreme events. Multilevel modeling is based on the consideration of stochastic and dynamic-stochastic time series of climatic, production-economic, environmental, biological, and other characteristics in the form of hierarchical structures with separate descriptions of sequences of local minima, local maxima, and intermediate values.

The proposed multi-level trend models are used to forecast crop yields and yields, labor costs for producing various products, and to assess the dynamics of climate trends and cost indicators. In addition to identifying average, favorable, and unfavorable trends, the probabilistic variability of favorable and unfavorable events can be identified. This allows for obtaining comprehensive information on the properties of various characteristics, assessing their dynamic and random components.

Since crop yields are linked to meteorological factors, multi-level trend models are used to distinguish between low and high crop yields. This allows for assessing the similarities and differences in meteorological conditions during the identified extreme years.

The developed multi-level parametric mathematical programming problem can be applied in various sectors of agriculture: crop production, a combination of livestock and crop production, processing, a combination of production and processing, and a combination of production, processing, and marketing. Specifically, this problem is used to determine optimal solutions for food production and processing under various producer operating conditions using two-stage modeling. In the first stage, production volumes are calculated under average, favorable, and unfavorable conditions, while in the second stage, the volumes of processed products are calculated based on the raw materials received. Multilevel, multistage modeling can be used for planning under favorable and unfavorable extreme events, based on the calculated probability intervals of these events.

A two-stage multilevel parametric problem can be supplemented by a stage of sales of manufactured and processed products. In this case, the constraints characterizing product sales are related to the volumes of manufactured and processed products, which are determined as unknown values in the first and second stages of modeling.

The presence of multilevel trends in forage crop yield increases determines the possibilities for modeling the numbers of farm animals of different species in pasture-based livestock farming, with and without accounting for agricultural land degradation.

Multilevel models are used to plan agricultural production at various levels of aggregation, enabling the optimization of output at an agricultural enterprise, within a group of enterprises, within a municipal district, at the level of an agricultural landscape region, agricultural zone, and region. Planning scenarios under potential risks are also considered.

FASTEST COLLISION-FREE ROUTES IN AN ENVIRONMENT WITH DYNAMIC OBSTACLES

Kazakov Aleksandr[1], Lempert Anna[1], Tran Thai Viet[2]

[1]Matrosov Institute for System Dynamics and Control Theory of SB RAS
[2]Irkutsk National Research Technical University
Irkutsk, Russia
ttviet@ictu.edu.vn

The problem of constructing time-optimal routes for one or several moving objects in a dynamic environment with obstacles is NP-hard and has a wide range of applications, from the coordination of unmanned aerial vehicles and service robots to traffic simulation. Classical methods (e.g., A*) are effective in static environments [1], but dynamic changes require adaptive routing strategies.

Problem Statement. Let we are given a bounded 2D domain X, along with a function $0 \le f(t,x,y) \le \beta$, which defines the instantaneous speed at point $(x,y) \in X$ at time t, and a set of barriers $H(t) = \{(x,y): f(t,x,y) = 0)\}$. For n moving objects $\Theta = \{O_1, O_2, ..., O_n\}$, find trajectories $P_i(t)$, $P_i(0) = S_i$, $P_i(T_i) = F_i$, for some $T_i > 0$, subject to collision-free conditions with barriers and other objects, considering a safety distance Δ, while minimizing the total travel time.

Solution Method. The proposed method is based on an optical-geometric analogy, in which an object's trajectory is identified with a light ray in an inhomogeneous medium [3]. A light wave propagation algorithm is used to calculate travel time between two points, forming a field of arrival times. A route construction algorithm (OGM) constructs a route for a single object considering the predicted positions of moving obstacles. A multi-route coordination algorithm (OGMN) handles systems with multiple moving objects, requiring not only static obstacle avoidance but also preventing mutual collisions. The key idea is iterative trajectory planning for each object over successive time intervals, where all other objects are treated as dynamic obstacles.

Results. We compare the proposed OGMN algorithm with CFM, PRIMAL, and DHC [2]. We consider two environments of size 40×40 and 80×80, where barrier density is 30%. The number of moving objects was 4, 8, 16, 32, and 64; their start and target points are random. The OGMN method shows a higher probability of reaching the target point than PRIMAL and DHC. For 64 objects in a 40×40 environment, the efficiency of the OGMN is 65.2%, while PRIMAL and DHC achieved 20% and 49%, respectively. In the 80×80 environment, the OGMN showed 96.5% efficiency, while PRIMAL and DHC achieved 12% and 74.5%. Moreover, the OGMN method finds routes with lower average completion time than PRIMAL and DHC.

References

1. Matthews J. Basic A* pathfinding made simple. AI Game Programming Wisdom // Boston: Charles River Media, 2002. P. 105–113.
2. Sartoretti G., Kerr J., Shi Y., Wagner G., Kumar T.K.S, Koenig S., Choset H. PRIMAL: Pathfinding via Reinforcement and Imitation Multi-Agent Learning // In IEEE Robotics and Automation Letters. 2018. V.3, No.4. P. 2378 – 2385.
3. Kazakov A.L., Lempert A.A. An approach to optimization in transport logistics // Automation and Remote Control. 2011. V. 72, No. 7. P. 1398 – 1404.

MACHINE LEARNING IN PARAMETRIC OPTIMIZATION OF AGRICULTURAL PRODUCTION VOLUMES

Klimov Egor, Ivanyo Yaroslav

Irkutsk State Agrarian University named after A.A. Ezhevsky
Molodezhny, Russia
rector@igsha.ru

Effective management of agricultural production under risky farming conditions requires the application of modern digital technologies. One of the key challenges is the high variability of crop yields driven by climatic factors. Recently, considerable attention has been devoted to artificial intelligence methods and algorithms for solving forecasting and planning tasks in agrarian production. The relevance of this research, related to the management of agricultural sectors, lies in developing tools that not only adequately forecast production and economic indicators using machine learning algorithms but also leverage these data for making informed decisions on resource use optimization.

This paper describes an algorithm for forecasting the yields of grain crops, potatoes, and vegetables for 2025–2030 using gradient boosting (LightGBM, CatBoost) and the Temporal Fusion Transformer (TFT) neural network architecture. The study was conducted on time series of crop yields from the Irkutsk District for 1996–2024. For the gradient boosting models, lag and rolling features based on preceding yield values were additionally computed.

The selection of the best model was carried out according to the principle of sliding validation. Information for 2020–2024. were used as a sample, and training was conducted on data from previous years. The average absolute error (MAE) was chosen as the main metric; the symmetric mean relative error (sMAPE) was additionally calculated. For each agricultural crop, the model with the smallest error values was selected.

The result of the application of the machine learning algorithm is the forecast values of the yield of agricultural crops for 2025–2030. These values were compared with the productivity obtained with the help of trends and the probability distribution of this characteristic for stochastic time series.

The obtained yield forecasts serve as the basis for solving the problem of parametric optimization, focused on the maximization of revenue. Meteorological factors (air temperature and precipitation during the growing season) are used as a parameter or parameters of the model.

For each year of the period 2025–2030, a linear programming problem was solved taking into account projected agricultural yields. Arable land area and labor force availability remained constant. The dynamics of the value of agricultural output were modeled.

Thus, an approach to forecasting crop yields, including gradient boosting algorithms (LightGBM, CatBoost) and the Temporal Fusion Transformer neural network architecture, was implemented. For each of the six crops (wheat, barley, oats, potatoes, cabbage, beets, potatoes), a model with the lowest error was automatically selected on the validation sample of 2020–2024.

Based on the dependence of crop yields on meteorological factors, a parametric programming model has been formulated and implemented for optimizing agricultural production volumes using data from the Irkutsk District. The model can be applied to an individual agricultural enterprise, an agrolandscape district, or an agricultural zone. A comparative analysis of the obtained results with the optimal solutions of the parametric problem, in which time was used as the parameter, has been conducted.

AN APPROACH TO THE AUTOMATION OF COGNITIVE MODEL CONSTRUCTION: TECHNIQUE, ALGORITHMS, AND SOFTWARE

Kozlov Maksim, Mikheev Aleksei

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia

Modern optimization models of large-scale energy systems, including those implemented in frameworks such as PyPSA [1], have reached a high level of technical detail. Yet their outputs often remain difficult to interpret for decision-makers, particularly in planning settings characterised by deep uncertainty. In practice, the analysis of how modelling assumptions affect results is frequently limited to local sensitivity checks or correlation-based assessments and therefore does not provide a reproducible, end-to-end explanation of which inputs drive the variability of key performance indicators and what the expected effect of changing them is.

This work proposes an integrated approach that combines computational experimentation, global sensitivity analysis [2], and causal effect estimation. The methodology consists of three main components: (1) repeated optimization runs under systematic variation of input parameters sampled from prescribed distributions; (2) Sobol'-based global sensitivity analysis to quantify each parameter's contribution to the variance of selected performance metrics [3]; and (3) estimation of the effects of input perturbations in the formalism of Pearl's structural causal models [4] using the DoWhy library [5], with explicit statements of identifiability assumptions and robustness checks. The results are consolidated into a cognitive map, represented as a directed graph in which nodes correspond to parameters and indicators, while edges encode stable relationships that are interpreted as causal links under the stated assumptions, together with effect sizes and uncertainty measures.

A Python-based software prototype has been developed to implement the proposed pipeline, from parameter sampling and orchestration of optimization experiments to cognitive map construction and visualization. The approach is demonstrated on an Italian power system model: the CO_2 emissions tax is shown to exert a pronounced negative impact on profit, its contribution to the variability of the performance metric is high according to global sensitivity analysis, and the estimated perturbation effect remains statistically robust within the adopted causal model.

Overall, the proposed methodology provides a reproducible and interpretable link between a complex optimization model of energy system and the needs of the analyst, and it can serve as a foundation for decision support systems in energy planning.

References

1. Hörsch J. et al. PyPSA: Python for Power System Analysis // Journal of Open Source Software. 2018. Vol. 3, № 32. P. 783
2. Saltelli A. et al. Global Sensitivity Analysis: The Primer. Wiley, 2008.
3. Sobol' I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates // Mathematics and Computers in Simulation. 2001. Vol. 55, № 1–3. P. 271–280.
4. Pearl J. Causality. Cambridge University Press, 2009.
5. DoWhy: A Python library for causal inference. Microsoft Research. URL: https://github.com/py-why/dowhy (дата обращения: 2026).

ARCHITECTURE OF THE DIGITAL PLATFORM OF THE KNOWLEDGE ECOSYSTEM IN ENERGY

Kuzmin Vladimir

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
kuzmin_vr@isem.irk.ru

The Department of Artificial Intelligence Systems in Energy is currently developing a Digital Platform of the Knowledge Ecosystem in Energy. A Digital Platform is defined as a system of tools that supports the use of digital processes, resources, and services by a significant number of digital ecosystem entities and enables their seamless interaction. That is, the concept of a Digital Platform for a Knowledge Ecosystem includes not only computational and mathematical models, but also databases and knowledge bases, information and artificial intelligent subsystems, a visualization subsystem, and the necessary services. Figure 1 shows the general architecture of a Digital Platform of the Knowledge Ecosystem in the Energy.

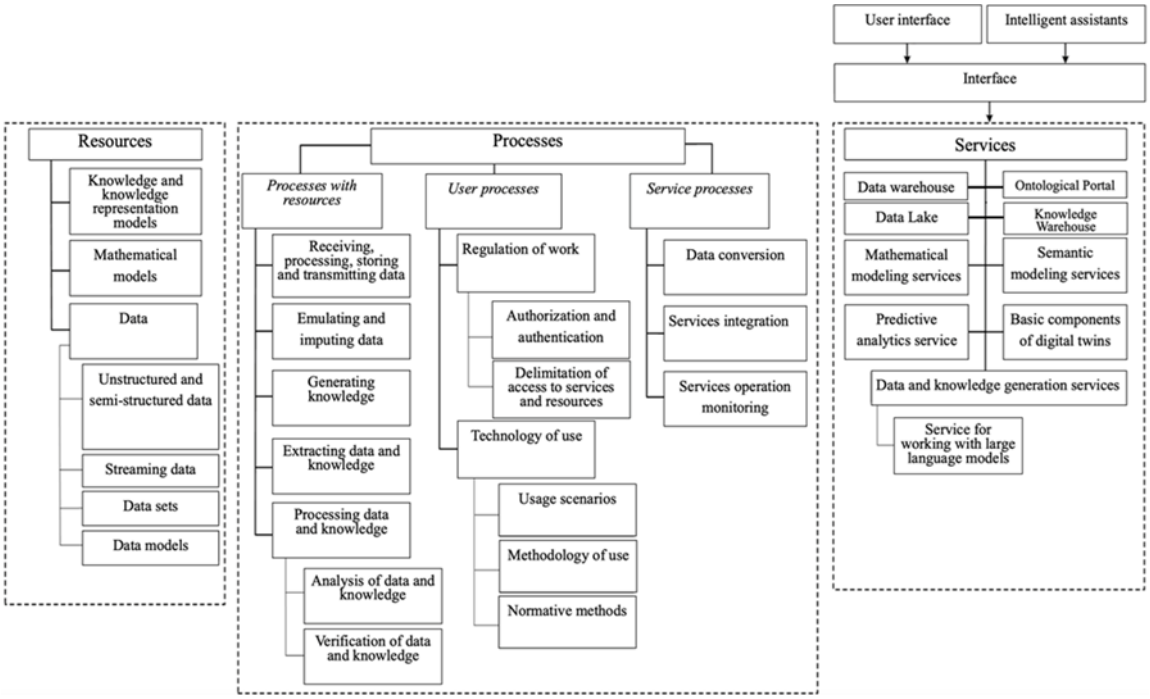


Fig. 1. General architecture of the Digital Platform of the Knowledge Ecosystem in the Energy

The report will consider the specification of the proposed architecture, describe the main services and components used in them, and also consider examples of component integration.

APPLICATION OF THE INTEC SOFTWARE PACKAGE IN STUDIES OF ENERGY SYSTEM RESILIENCE

Mamedov Timur

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
mamedowtymur@yandex.ru

Studies of energy system resilience are aimed at analyzing and substantiating the development directions of the fuel and energy complex (FEC), taking into account the requirements of energy security, environmental sustainability, and quality of life of the population [1]. Under modern conditions, such studies are characterized by high complexity and a multivariant nature, which is caused by the need to consider various scenarios for the development of energy systems, including the impact of external threats, resource constraints, and possible management measures.

At the Melentiev Energy Systems Institute of the Siberian Branch of the Russian Academy of Sciences, the INTEC software package has been developed to support forecasting studies of the FEC development and strategic decision-making in the energy sector [2]. The software package implements methods of mathematical modeling and enables computational experiments based on techno-economic models of energy systems. The models are based on the optimization of fuel and energy balances formulated as a linear programming problem, where the objective function minimizes the total discounted costs of the energy system operation and the losses associated with energy resource shortages among consumers [3].

A distinctive feature of modern studies of energy system resilience is the need to take into account not only quantitative parameters of the functioning of the FEC but also qualitative factors related to energy security threats and management actions. In this regard, in recent years an approach based on the integration of mathematical and intelligent modeling, including the use of ontologies and cognitive maps, has been actively developed.

Within this approach, ontologies are used to structure domain knowledge and formalize relationships between the elements of the techno-economic model. Based on ontological models, cognitive maps are constructed to represent causal relationships between the factors of energy system development, threats, and compensatory measures. Such maps serve as a tool for forming computational scenarios and visualizing modeling results [2].

The INTEC-SAW software package implements the proposed technology for integrating semantic and mathematical models. Its functionality includes preparation of the basic variant of the FEC model, formation of development scenarios by modifying parameters of variables and constraints, optimization of models for various development options, and interpretation of results in the form of balance tables and cognitive maps.

The use of the INTEC software package in studies of energy system resilience makes it possible to conduct computational experiments aimed at assessing the consequences of various threats and the effectiveness of compensatory measures. In particular, this toolkit is used to analyze the influence of natural, technological, and economic factors on the functioning of energy systems. The results of modeling can be used to substantiate strategies for the development of the FEC and to develop measures aimed at ensuring energy security.

Acknowledgements. The research was carried out within the framework of the state assignment project (No. FWEU-2021-0007) of the Fundamental Research Program of the Russian Federation for 2021–2030.

References

1. Massel L. V. Digital platform for supporting strategic decision-making for the sustainable development of Russia’s energy sector. In: Sustainable Development of the Energy Sector of the

Republic of Belarus: State and Prospects, Proceedings of the III International Scientific Conference, Minsk, October 1–4, 2024. Minsk: Belaruskaya Navuka, 2025. pp. 44–53.
2. Massel A. G., Mamedov T. G. Integration of mathematical and cognitive modeling in studies of the development directions of the fuel and energy complex from the perspective of energy security. Information and Mathematical Technologies in Science and Management. 2025. No. 3(39). pp. 61–71. DOI: 10.25729/ESI.2025.39.3.006.
3. Pyatkova N. I., Mamedov T. G. Features of studying the influence of energy security threats on the reliability of energy supply to consumers under modern conditions. Information and Mathematical Technologies in Science and Management. 2025. No. 4(40). pp. 128–137. DOI: 10.25729/ESI.2025.40.4.010.

PROBLEMS OF DEVELOPMENT AND PROSPECTS FOR APPLICATION KNOWLEDGE ECOSYSTEMS IN ENERGY

Massel Liudmila

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
massel@isem.irk.ru

The staff of the department "Artificial Intelligence Systems in Energy" continues to develop a knowledge ecosystem in energy, which was proposed as an innovative approach to knowledge management [1] and is being developed on the basis of the IT infrastructure for systems research in energy, developed by the department earlier [2].

A knowledge ecosystem is an organization composed of various actors (developers and beneficiaries) united by a shared search for valuable knowledge, while simultaneously engaging in independent activities outside the knowledge ecosystem. Knowledge ecosystems include interconnected knowledge resources, databases, and experts.

The foundation of a knowledge ecosystem is a digital platform. This is understood as a system of tools that supports the use of digital processes, resources, and services by a significant number of digital ecosystem actors and enables their seamless interaction[1]. In other words, the concept of a digital platform for a knowledge ecosystem includes not only computer and mathematical models but also databases and knowledge bases, information and intelligent subsystems, a visualization subsystem, and the necessary services. The generalized architecture (meta-ontology) of a knowledge ecosystem is presented in Figure 1.

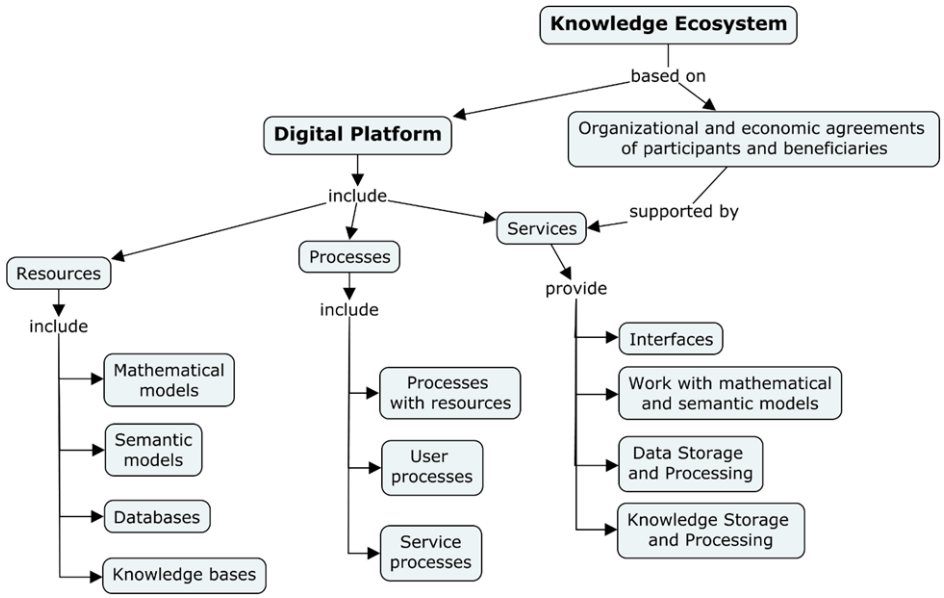


Fig. 1. The generalized architecture (meta-ontology) of a knowledge ecosystem

For the functioning and operation of a knowledge ecosystem, the digital platform must be supplemented with organizational and economic relationships between developers and beneficiaries.

Given the architecture of the digital platform, **the main methods for building a knowledge ecosystem are:**

- methods of mathematical and semantic modeling and predictive analytics;
- methods of knowledge extraction and representation, methods for designing data and knowledge warehouses;

[1] Decision of the Supreme Eurasian Economic Council of 11 October 2017 No. 12 "On the main directions of implementing the digital agenda of the Eurasian Economic Union until 2025"

- methods of knowledge generation and processing, including analysis and verification of data and knowledge;
- methods of constructing digital twins of energy facilities and emulating and imputing data for them;
- methods of integrating services and monitoring their operation;
- methods of knowledge mapping;
- methods of interface development, including the construction of intelligent assistants using large language models [3].

The main problems in creating knowledge ecosystems are:

1. This area has not gained much traction in Russia, despite the fact that, for example, only one foreign article alone cites 80 references on this topic. Russian literature displays inconsistent terminology: the terms "knowledge ecosystems," "knowledge management ecosystems," and "digital knowledge management infrastructure" are used as synonyms.

2. The closed nature of data and knowledge—it is declared that "corporate knowledge is the property of the corporation." Data is even more closed.

3. A lack of understanding of the problem: for example, we encountered the opinion of an RSF expert that "knowledge and energy systems are different scientific fields and cannot be considered in a joint scientific project." Furthermore, in some cases, resistance from specialists is possible, for example, regarding attempts to use tacit knowledge.

4. It is unclear how to build relationships between developers and beneficiaries. The lack of regulations (for example, in the area of digital twins) also poses an obstacle.

5. COMPLEXITY OF IMPLEMENTATION (multi-skilled specialists are required, with qualifications in subject areas, in the field of database and knowledge base design, software development, use of large language models, etc.).

Prospects for applying a knowledge ecosystem in the energy sector:

1. The proposed digital knowledge ecosystem platform can be used for various energy research projects, as it has a flexible architecture that allows for the integration of additional components.

In particular, the digital platform will enable the modeling of four energy infrastructure levels (unit, energy facility, energy system, and the country's fuel and energy complex) and the use of software tools and technologies for conducting research on these levels.

2. The knowledge ecosystem is planned to support research on resilience energy systems, integrating research on energy security, environmental resilience, and quality of life.

3. The developers believe the primary application of the digital platform is intelligentl support for strategic decision-making regarding the development of Russia's energy sector.

References

1. Massel L.V. Knowledge ecosystem as development and specialization of the digital ecosystem. Proceedings of the International Scientific and Technical Congress "Intelligent Systems and Information Technologies-2023". Scientific publication in 2 volumes. V.2. – Taganrog: Publisher Stupin S.A., 2023. – Pp. 155-164.
2. Massel L.V., Massel A.G. Building a knowledge ecosystem based on the IT infrastructure of system research in energy / Bulletin of the University of Yugra, 2023. – No. 4. – P. 78-87.
3. Massel L.V., Massel A.G., Kuzmin V.R. Methods for constructing a knowledge ecosystem using the energy sector as an example // XXII National Conference on Artificial Intelligence with International Participation, KII-2025 (St. Petersburg, October 6-10, 2025). Conference Proceedings. In 3 volumes. Vol. 1. – St. Petersburg: Publishing House of St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2025. – Pp. 59-68/

ARCHITECTURE OF THE DIGITAL TWIN VISUALIZATION COMPONENT USED IN THE KNOWLEDGE ECOSYSTEM

Massel Liudmila, Shchukin Nikita

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
niksha14@mail.ru

This report presents the interim results of the research and development of the architecture of a digital twin visualization component operating within a knowledge ecosystem. Given the rapid development of the energy sector and the increasing complexity of energy facility design, the development of digital twins operating within an integrated knowledge ecosystem is particularly relevant.

Visualization issues in digital twins are particularly important, as visualization tools provide the interface between complex mathematical calculations and decision makers. The role of visualization tools in the effective functioning of digital twins is determined by the need to clearly present modeling and optimization results for operational analysis and management.

This report proposes an architecture for a digital twin visualization component aimed at solving two interrelated problems: the problem of finding the optimal location for a power facility and the problem of simulating its operating options.

The visualization component in the optimal object placement problem displays a digital terrain map and demonstrates the optimization results. This allows the designer not only to obtain numerical parameter values but also to visually evaluate the configuration of the future object.

The component's second functional task is to visualize the simulation modeling of the object's operation at a selected location. Here, the visualization component performs a different function: it must display the dynamic processes occurring in the system. The visualization architecture in this case must support an "as if" mode, allowing for observation of the digital twin's behavior under various scenarios.

Thus, the developed visualization component architecture significantly increases the transparency of digital twin models for decision-makers. The integration of semantic technologies and modern computer graphics tools within the knowledge ecosystem improves the quality of design decisions and the efficiency of management.

Acknowledgments. The research was carried out under State Assignment Project (FWEU-2026-0008) of the Fundamental Research Program of Russian Federation.

MODELING AND OPTIMIZATION OF SUSTAINABLE ENERGY SYSTEMS: PROBLEMS, APPROACHES, AND PARADOXES

Mikheev Aleksei

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
mikheev@isem.irk.ru

The paper examines methodological challenges in modeling and optimizing sustainable energy systems under the energy transition, increasing shares of variable renewables, deepening sector coupling, and stricter requirements for reliability, environmental admissibility, and social acceptability. It is argued that the classical cost-minimization paradigm is no longer sufficient for analyzing long-term transition pathways. Contemporary models must simultaneously address multi-criteria objectives, uncertainty, hierarchical decision structures, tractability constraints, and the interpretability of results, while consistently representing the temporal, spatial, technological, sectoral, and institutional dimensions of the system [1], [3].

Energy system modeling is therefore interpreted not as a search for a single optimum, but as an exploration of feasible and near-optimal solution spaces that differ in technology mix, risk exposure, reserve requirements, and resilience under stress scenarios. Special attention is paid to the specifics of sustainable energy systems, where costs and emissions must be assessed jointly with reliability, flexibility, energy affordability, intersectoral linkages, and life-cycle impacts. The paper argues that the central methodological tension lies between model fidelity and tractability: increasing temporal, spatial, and scenario detail rapidly leads to very large-scale optimization problems that, in practice, require decomposition, aggregation, surrogate models, and hybrid computational schemes combining optimization, simulation, and ML-based components.

Particular attention is devoted to modeling paradoxes, including aggregation and scale effects, uncertainty, renewable integration, network interactions, and efficiency-related counterintuitive effects, in which locally improving measures may produce systemically fragile or even less sustainable outcomes. In particular, the choice of temporal aggregation method may significantly affect investment decisions and the estimation of system costs; similarly, rebound effects and Braess-type phenomena show that higher efficiency or additional infrastructure do not automatically improve overall system performance [2], [4], [5]. The paper substantiates the need for stress validation, multi-scale analysis, MGA, MCDA, and human-in-the-loop procedures in order to derive more robust, transparent, and interpretable strategies for the development of critical energy infrastructure [3].

References

1. Ringkjøb H.-K., Haugan P. M., Solbrekke I. M. A review of modelling tools for energy and electricity systems with large shares of variable renewables // Renewable and Sustainable Energy Reviews. 2018. Vol. 96. P. 440-459. DOI: 10.1016/j.rser.2018.08.002.
2. Kotzur L., Markewitz P., Robinius M., Stolten D. Impact of different time series aggregation methods on optimal energy system design // Renewable Energy. 2018. Vol. 117. P. 474-487. DOI: 10.1016/j.renene.2017.10.017.
3. Cinelli M., Burgherr P., Kadziński M., Słowiński R. Proper and improper uses of MCDA methods in energy systems analysis // Decision Support Systems. 2022. Vol. 163. Art. 113848. DOI: 10.1016/j.dss.2022.113848.
4. Özsoy T. The “energy rebound effect” within the framework of environmental sustainability // WIREs Energy and Environment. 2024. Vol. 13, No. 2. Art. e517. DOI: 10.1002/wene.517.
5. Hartmann B., Ódor G., Benedek K., Papp I., Cirunay M.T. Quantitative comparison of power grid reinforcements // Chaos, Solitons & Fractals. 2025. Vol. 200, Part 3. Art. 117095. DOI: 10.1016/j.chaos.2025.117095.

A METHOD FOR INCREASING THE DIGITAL MATURITY OF TERRITORIAL SAFETY MANAGEMENT

Nicheporchuk Valerii

Institute of Computational Modeling SB RAS
Krasnoyarsk, Russia
valera@icm.krasn.ru

In this research, we suggest a methodology for digital transformation of territorial management. For described of the business processes for solving safety tasks with the support of intelligent tools used standards designing of information systems. It is possible to build adaptive and self-learning systems for management complex dynamic objects. The assessment of the statement and development prospects of such objects are difficult to predict.

The method has not goal for greater accuracy of computational modeling procedures, an increase in the range of indicators, and the formation of large time series. We solving the problem of improving the quality of decisions made by under qualified specialists and the incompleteness of the in conditions incomplete regulatory base. Structuring information resources allows for logical linking of strategic and operations decisions. The form of presentation of decisions ensures the replication of the experience of effective management through cyclical work with cognitive representations of the algorithmization of reasoning. [1]. Descriptions of real situations are transformed into standard solution templates that are used to provide information support in situations with similar conditions.

A taxonomy of management tasks is needed to specify safety problems. The classification of tasks includes basic risks [2], other characteristics of objects and processes. The assessment of the "atomicity" of a task is determined by experts. This is necessary for a compact description of solutions and their unambiguous interpretation. A generated list of tasks that can be ranked by implementation priority, resource availability, and used to control the completeness of strategic planning documents.

Multiple repetition of the cycle of forming management decisions is used as a justification for adjusting regulatory documents. The use of intelligent systems trained on real-world problems will reduce the requirements for the qualifications of decision-makers and identify bottlenecks and management errors.

Acknowledgements. The work was carried out within the framework of the project of the state assignment of the Federal Research Center KSC SB RAS of the fundamental research program of the Russian Federation (reg. FWES-2021-0004).

References

1. Nicheporchuk V.V., Kobyzhakova S.V. Digitalization Technology for Territorial Security Management // Information and Mathematical Technologies in Science and Management, 2025. No. 1 (37). Pp. 104-113. DOI: 10.25729/ESI.2025.37.1.010.
2. Moskvichev Vladimir V. Natural and technogenic safety of Siberian regions: a retrospective of research and new challenges // Computational Technologies, 2023. Vol. 28, no. 3, pp. 25–40. DOI:10.25743/ICT.2023.28.3.003

MODELING OF UNBALANCED OPERATING MODES OF LOW-VOLTAGE ELECTRIC NETWORKS

Ntsama Ndi Bernard Josselin, Naumov Igor

Irkutsk State Agrarian University named after A.A. Ezhevsky Molodezhny, Russia

The operating modes of existing distribution electric networks of a 0.4 kV voltage level are objectively unbalanced. This imbalance is due to both the asymmetry of the three-phase voltage system due to changes in the symmetrical components of the currents and voltages negative and zero sequences of the fundamental frequency, and the non-sinusoidal nature of this system due to changes of an additional sequences of positive, negative and zero components conditioned an emission of the currents and voltages higher harmonic components.

Indicators characterizing the level of asymmetry and harmonic distortion of the power supply voltage are established by state standards of various countries around the world. The indicators characterizing the generalized unbalanced mode of a low-voltage system operation have not been established by the standards and are being considered for the first time. Such criteria for assessing of the electrical network unbalance level include generalized coefficients of the voltage unbalance of positive, negative and zero sequences up to the 40th harmonic, the total coefficient of voltage unbalance, voltage deviation (this indicator is established by standards) and the coefficient of an active power losses increase.

It has been established that the most effective balancing means in an operating mode of low-voltage electrical network is a universal balancing device (BD) developed at the Department of Power Supply named A.A. Yezhevsky IrGAU, the parameters of which automatically change as a function of the imbalance changing level in the studied electrical network.

We are currently offering a simulation model of an electrical network developed in the MATLAB Simulink program, feeding an unbalanced load of consumers creating various levels of unbalanced electrical consumption. The change in the established unbalance indicators is investigated in the absence of a control unit of the studied model, as well as during simulation integration into the studied electrical network of this device. It is proved that the proposed balancing device model effectively reduces the unbalance indicators, which leads to a significant increase of the power quality and a reduction in its losses.

TELEMETRY DATA PREPARATION TECHNOLOGY FOR TRAINING HYBRID NEURAL NETWORKS IN PREDICTIVE ANALYTICS

Shtyrkin Philip

Irkutsk National Research Technical University
Irkutsk, Russia
shtyrkinfa@ex.istu.edu

Modern predictive maintenance systems for mining equipment are limited in their feasibility without the usage of machine learning methods. In this study, while developing a system for predicting dump truck failures, a problem appeared: the source telemetry data in the Vehicle Health Monitoring System (VHMS) format [1] proved unsuitable for direct training.

Primary analysis revealed four key aspects that made the source data unsuitable for training an artificial neural network (ANN):

1. Unstructured format
2. Asynchronicity and missing data
3. Multi-scale and statistical heterogeneity
4. Critical class imbalance

To overcome these barriers, the following step-by-step data transformation technology was implemented, consisting of three key stages.

The first step was dataset denormalization. The table was transposed to improve interpretation. Linear interpolation was used to eliminate gaps due to asynchronous recording.

To normalize all features, the Z-score normalization method was chosen. The Z-score transformation (standardization) allowed the data to be distributed with a mean of 0 and a standard deviation of 1, thus addressing the multi-scale data issue.

The data volume after interpolation remained redundant (hundreds of thousands rows), making training resource-intensive and containing high-frequency noise. The solution was the use of the Fast Fourier Transform (FFT) [2, 3].

For each time window, the signal spectrum was calculated. The FFT allowed the signal to be converted from the time domain to the frequency domain, thereby removing high-frequency noise components and preserving only the significant spectral descriptors. The key result was radical data compression: the sample size was reduced from 550,000 time samples to 1020 spectral images. At the same time, the diagnostic value was preserved, since many failures manifest themselves in changes in the frequency characteristics of component operation, rather than in instantaneous values.

Thus, the computationally intensive task of training an ANN for failure prediction was solved not only through the development hybrid neural network architecture (long short-term memory (LSTM) + classical perceptron), but above all through a deep and scientifically sound transformation of the source data. The proposed technology made it possible to transform an unstructured, noisy, and unbalanced telemetry stream into a compact dataset suitable for training an ANN in a predictive analytics task.

References

1. Chen, S., & Zhu, M. (2010). Design and Implementation of Health Monitoring System on Electric Wheel Truck. 2010 International Symposium on Intelligence Information Processing and Trusted Computing. doi:10.1109/iptc.2010.149
2. Korovkin, N. V. On the Applicability of the Fast Fourier Transform for Harmonic Analysis of Nonsinusoidal Currents and Voltages / N. V. Korovkin, S. S. Gritsutenko // Bulletin of the Russian Academy of Sciences. Power Engineering. - 2017. - No. 2. - Pp. 73-86. - EDN WFFWIZ.
3. Yakimov, V. N. Reconstruction of a Discrete Signal Sequence Based on a Moving Average Model and Estimation of the Correlation Relationship of Samples in Forward and Backward Forecasting / V. N. Yakimov // Computer Science and Automation. - 2024. - Vol. 23, No. 3. - Pp. 766-800. – DOI 10.15622/ia.23.3.5. – EDN PQCDDK.

THE "FILLER & POINTER" APPROACH FOR IMPROVED SATELLITE-BASED SUPERVISION OF LINEAR CLEAR-CUTS IN ENERGY, TRANSPORT, AND OIL/GAS INFRASTRUCTURE EXPANSION

Sokolkov Oleg

Yugra State University
Khanty-Mansiysk, Russia

This study addresses the challenge of automating the detection of forest clearances using Earth observation data within the framework of effective enforcement of forest protection laws. Given that manual interpretation of satellite imagery suffers from low efficiency due to human limitations like fatigue and reduced focus over extended periods of repetitive tasks, there has been an increasing adoption of automated monitoring systems utilizing neural network techniques, driven by advancements in artificial intelligence technology. These hybrid approaches combine human interpretative skills with the computational capacity of neural networks, achieving both precision and scalability when analyzing vast amounts of satellite data, thereby enhancing efforts to safeguard forest ecosystems. Despite notable successes in applying neural networks to this domain, real-world implementations face significant obstacles related to inadequate detection rates for specific types of deforested zones intended for road construction, power line installation, or pipeline laying. To address these shortcomings, we propose a novel approach to neural semantic segmentation, which involves splitting the segmentation task into two distinct subtasks: filling and pointing, accompanied by a scheme for aggregating the outputs of these subtasks into a unified result. We validate our approach using a dataset derived from forest clearance incidents recorded across the Khanty-Mansi Autonomous Okrug – Yugra region between 2018 and 2022. The data consists of labeled images from the Sentinel-2 satellite and covers June–September season. The proposed approach improved the accuracy of linear clear-cuts detection in terms of the F1-score by 5.15% without compromising the recognition quality of other types of forest clearings.

PRACTICAL APPLICATION OF SWARM-BASED GLOBAL OPTIMIZATION ALGORITHMS IN ATOMIC AND MOLECULAR MODELING

Sorokovikov Pavel

Matrosov Institute for System Dynamics and Control Theory of SB RAS
Irkutsk, Russia
pavel@sorokovikov.ru

There has been a steady increase in interest in the formation of ultrafine structures of atomic and molecular clusters. The primary objective of such studies is to determine a cluster configuration corresponding to the minimum potential energy. Most problems related to the search for low-energy states of atomic and molecular clusters are formulated as the minimization of nonconvex functions and are characterized by an extremely rapid growth in the number of local optima as the dimensionality of the variable space increases. The problem of finding the global minimum of a multimodal objective function remains one of the most challenging and relevant tasks in both theoretical optimization and practical applications. This work considers the minimization of a multimodal function represented by the Morse potential [1].

In recent years, significant attention has been devoted to the development of numerical methods for global optimization. To date, a large number of search-based algorithms have been proposed, differing in their computational characteristics and application domains. Among them, metaheuristic methods inspired by natural processes occupy a special place, including swarm intelligence algorithms [2, 3]. Their growing popularity is driven by advances in computational power and the need to efficiently solve high-dimensional optimization problems.

This paper presents a numerical study of more than 20 modern swarm-based global optimization algorithms applied to the problem of minimizing the Morse potential. The swarm intelligence methods considered include the Firefly Algorithm, Coyote Optimization Algorithm, Gorilla Troops Optimizer, Tuna Swarm Algorithm, Ant Lion Optimizer, and others [2, 3]. The algorithms were tested on atomic and molecular clusters containing up to 80 atoms (240 variables). Computational results obtained by each swarm algorithm operating independently were compared with those achieved when combined with the L-BFGS local search method. The results of numerical experiments demonstrating the effectiveness of the applied algorithms are presented.

Acknowledgements. The work was carried out at the expense of a grant from the Ministry of Science and Higher Education of the Russian Federation, the project No.: 126021217177-7.

References

1. Doye J.P.K., Wales D.J. Structural consequences of the range of the interatomic potential a menagerie of clusters // J. Chem. Soc. Faraday Trans. 1997. Vol. 93(24). P. 4233–4243.
2. Jakšić Z., Devi S., Jakšić O., Guha K. A comprehensive review of bio-inspired optimization algorithms including applications in microelectronics and nanophotonics // Biomimetics. 2023. Vol. 8(3). P. 278.
3. Devika G., Karegowda A.G. Bio-inspired optimization: algorithm, analysis and scope of application // Swarm Intelligence-Recent Advances and Current Applications. 2023. P. 1–37.

ARCHITECTURE OF A DATA LAKE FOR THE ENERGY KNOWLEDGE ECOSYSTEM

Tsybikov Aleksei

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
tsibikow@mail.ru

The architecture of a Data Lake within an energy knowledge ecosystem serves as a fundamental technological layer that enables the consolidation and systematization of dissimilar industry data. In the energy sector, data is generated by numerous sources, ranging from technological monitoring and control systems to corporate information systems, regulatory documentation, and analytical reports. Traditional data storage systems, primarily designed for structured information, are often insufficiently flexible to handle such a diversity of formats and volumes of data. The use of a Data Lake makes it possible to store data in its original form and create a unified storage environment that supports further processing, analysis, and knowledge extraction.

A key feature of Data Lake architecture is the multi-layered organization of data storage and processing. Typically, several logical zones are distinguished: the raw data zone, the processed and normalized data zone, and the curated analytics zone. At the first level, data is accumulated in its original format without significant transformation, ensuring the possibility of data reuse and the reproducibility of analytical processes. At subsequent levels, data cleaning, standardization, and integration from multiple sources are performed, after which specialized datasets are created for analytics, machine learning, and decision-support tasks.

Within the platform architecture, the data integration layer plays an important role by ensuring the ingestion of data into the Data Lake from various information systems. This involves both batch data loading and streaming data transmission, which is particularly relevant for energy systems that generate continuous streams of telemetry and technological parameters. The integration layer provides data extraction, initial transformation, and routing, as well as monitoring of data quality and consistency. This approach enables the integration of technological, operational, economic, and external data within a unified storage infrastructure.

Thus, the Data Lake architecture forms the foundation for building an energy knowledge platform by providing scalable data storage, support for multiple types of information, and the possibility of subsequent intelligent processing. The presence of a unified centralized data repository improves the efficiency of analytics, facilitates the development of digital services, and creates the conditions for applying machine learning and advanced data analytics methods in the management of energy systems. As a result, the Data Lake becomes a key element of the digital infrastructure of modern energy systems and an important instrument for the formation of industry knowledge.

Acknowledgements. The research was carried out under State Assignment Project (FWEU-2026-0008) of the Fundamental Research Program of Russian Federation.

COMPOSITION AND META DESCRIPTIONS OF DATALAKE COMPONENTS FOR ENERGY RESEARCH

Vorozhtsova Tatiana

Melentiev Energy Systems Institute SB RAS
Irkutsk, Russia
tnn@isem.irk.ru

Within the framework of the ecosystem of knowledge in the energy sector being developed at ISEM, a system of accumulation, storage, updating, and verification of data and knowledge is being designed and implemented [1]. The knowledge ecosystem is proposed as an innovative approach to knowledge management and is being developed based on the IT infrastructure of systems research in the energy sector. Knowledge ecosystems include interconnected knowledge resources, databases, knowledge bases, knowledge repositories, and expert knowledge.

The topic of energy systems research requires the analysis of a large amount of data and knowledge. Knowledge bases and data warehouses are an important component of the knowledge ecosystem. Modern means of working with such information are various forms of data warehouses. The DataLake architecture is proposed for storing and accumulating data in the knowledge ecosystem being developed at ISEM [2]. A data lake is a storage that receives both structured and unstructured data using ELT (Extract Load Transform) technology. This allows you to store and analyze a wider range of data, including text documents in various formats, images, sensor readings, etc. The data lake allows you to extract knowledge from non-obvious sources of information.

At the initial stage of the development of DataLake for energy research, it is proposed to consider such basic components as documents, databases, and data tables designed for analyzing big data in their original format.

Data Lake can support structured data from relational databases or data in other traditional formats, semi-structured data from applications and web servers, as well as unstructured data including text documents of various formats, images, videos, audio, presentations, emails, and others.

When uploading to the lake, each data set is assigned key tags (metadata) – data type, data source, time of receipt, data format [3]. In the next step, the metadata describes the components depending on the type of data. For documents, such metadata is the document type, title, creation date, and source. Depending on the type of document, the following descriptions are added. The following metadata is important for scientific literature – publications, articles, monographs, textbooks, dictionaries, scientific reports, scientific projects: author(s), keywords, abstract, edition, DOI, ISBN, etc. For technical documentation (technical reports, technical specifications, regulations, manuals, instructions, specifications), in addition to standard metadata, the status (approved, draft), access level, and others are taken into account.

The structure of meta descriptions (metadata) of DataLake components depends on the types of data (structured, poorly structured, unstructured), as well as on the purposes of using this metadata – for storage, for data processing, for analytics, for security, etc.

Acknowledgements. The work is carried out within the framework of the state assignment project "Methods of artificial intelligence and situational management in research and ensuring the sustainability of energy and environmental systems as interconnected critical infrastructures", No. FWEU-2026-0008. fundamental research program of the Russian Federation for 2021-2030.

References

1. Massel L.V. Ecosystem of knowledge as the development and specialization of the digital ecosystem. Proceedings of the International Scientific and Technical Congress "Intelligent Systems and Information Technologies-2023". Scientific edition in 2 volumes. Vol.2. Taganrog: Publisher Stupin S.A., 2023. pp. 155-164.

2. A new generation data lake: a 10–fold increase in productivity! URL: https://datafinder.ru/products/ozero-dannyh-novogo-pokoleniya-uvelichenie-proizvoditelnosti-v-10-raz?ysclid=mlrgxhjsnf127544927
3. Pegdwendé N. Sawadogo, Étienne Scholly, Cécile Favre, Éric Ferey, Sabine Loudcher, Jérôme Darmon. Metadata Systems for Data Lakes: Models and Features. URL: https://www.academia.edu/72192535/Metadata_Systems_for_Data_Lakes_Models_and_Features

CONTENT

Alekseev Roman, Lukyanov Nikita, Massel Liudmila	IMPLEMENTATION OF A NEURO-ASSISTANT BASED ON THE RAG-SYSTEM	2
Ananyev Leonid, Ivanyo Yaroslav	ALGORITHM FOR MODELING A MULTI-STAGE MULTI-LEVEL PROBLEM OF OPTIMIZING THE PRODUCTION AND PROCESSING OF AGRICULTURAL PRODUCTS	3
Bachurin Roman, Zakharova Irina	TIME SERIES CLUSTERING FOR THE ANALYSIS OF BIOMASS PYROLYSIS TG DATA	4
Chekan Mikhail	DIGITAL TECHNOLOGY OF AUTOMATION FOR AGENT MODELING OF MICROGRID INTERACTIONS	6
Enkhbat Ariunjargal, Ivanyo Yaroslav	MODELING THE POPULATION OF LIVESTOCK IN RELATION TO THE NATURAL-ECONOMIC ZONES OF MONGOLIA	8
Erzhenin Roman	ARCHITECTURE OF A MULTI-AGENT DISTRIBUTED CONTROL SYSTEM FOR THE FUNCTIONING OF AN ACCOUNTING INFORMATION SYSTEM	9
Fedotov Nikita	AUTOMATION OF REQUEST PROCESSING BASED ON AI AND STOCHASTIC MODELS	11
Galperova Elena, Galperov Vasilii	ARTIFICIAL INTELLIGENCE TECHNOLOGIES TO INCREASE THE ENERGY EFFICIENCY OF ECONOMIC SECTORS	12
Garashchenko Aleksandr, Tomin Nikita, Dreglea Alena, Sidorov Denis	GEOMAGNETIC INDUCED CURRENTS AND IONOSPHERIC DISTURBANCES AS RISK FACTORS FOR ENERGY SYSTEMS	13
Gilmutdinov Vladimir, Popov Maksim	ON DEFECTS OF WELDED JOINTS IN STEEL PENSTOCKS OF HYDROELECTRIC POWER PLANTS	14
Gornov Aleksandr, Anikin Anton, Lykhin Semen	FACTOR ANALYSIS OF RESPIRATORY TRACT INJURY RISKS USING THE SHEPARD OPERATOR	15
Govorkov Aleksei	THE MAIN DIRECTIONS OF QUALITY CONTROL OF PROJECT DOCUMENTATION IN AVIATION INDUSTRY	16
Ismagulov Milan	A HIERARCHICAL GRAPH-BASED MODEL OF DOCUMENT STRUCTURE FOR CONTROLLED LECTURE SUMMARY GENERATION	18
Ivanyo Yaroslav, Petrova Sofia	APPLICATIONS OF MULTILEVEL MODELING IN AGRICULTURAL PRODUCTION MANAGEMENT	21
Kazakov Aleksandr, Lempert Anna, Tran Thai Viet	FASTEST COLLISION-FREE ROUTES IN AN ENVIRONMENT WITH DYNAMIC OBSTACLES	22
Klimov Egor, Ivanyo Yaroslav	MACHINE LEARNING IN PARAMETRIC OPTIMIZATION OF AGRICULTURAL PRODUCTION VOLUMES	23
Kozlov Maksim, Mikheev Aleksei	AN APPROACH TO THE AUTOMATION OF COGNITIVE MODEL CONSTRUCTION: TECHNIQUE, ALGORITHMS, AND SOFTWARE	24
Kuzmin Vladimir	ARCHITECTURE OF THE DIGITAL PLATFORM OF THE KNOWLEDGE ECOSYSTEM IN ENERGY	25
Mamedov Timur	APPLICATION OF THE INTEC SOFTWARE PACKAGE IN STUDIES OF ENERGY SYSTEM RESILIENCE	26
Massel Liudmila	PROBLEMS OF DEVELOPMENT AND PROSPECTS FOR APPLICATION KNOWLEDGE ECOSYSTEMS IN ENERGY	28

Massel Liudmila, Shchukin Nikita	ARCHITECTURE OF THE DIGITAL TWIN VISUALIZATION COMPONENT USED IN THE KNOWLEDGE ECOSYSTEM	30
Mikheev Aleksei	MODELING AND OPTIMIZATION OF SUSTAINABLE ENERGY SYSTEMS: PROBLEMS, APPROACHES, AND PARADOXES	31
Nicheporchuk Valerii	A METHOD FOR INCREASING THE DIGITAL MATURITY OF TERRITORIAL SAFETY MANAGEMENT	32
Ntsama Ndi Bernard Josselin, Naumov Igor	MODELING OF UNBALANCED OPERATING MODES OF LOW-VOLTAGE ELECTRIC NETWORKS	33
Shtyrkin Philip	TELEMETRY DATA PREPARATION TECHNOLOGY FOR TRAINING HYBRID NEURAL NETWORKS IN PREDICTIVE ANALYTICS	34
Sokolkov Oleg	THE "FILLER & POINTER" APPROACH FOR IMPROVED SATELLITE-BASED SUPERVISION OF LINEAR CLEAR-CUTS IN ENERGY, TRANSPORT, AND OIL/GAS INFRASTRUCTURE EXPANSION	35
Sorokovikov Pavel	PRACTICAL APPLICATION OF SWARM-BASED GLOBAL OPTIMIZATION ALGORITHMS IN ATOMIC AND MOLECULAR MODELING	36
Tsybikov Aleksei	ARCHITECTURE OF A DATA LAKE FOR THE ENERGY KNOWLEDGE ECOSYSTEM	37
Vorozhtsova Tatiana	COMPOSITION AND META DESCRIPTIONS OF DATALAKE COMPONENTS FOR ENERGY RESEARCH	38

Proceeding of International Workshop:

"Critical Infrastructures in the Digital World"

(IWCI-2026)