

**Институт динамики систем и теории управления имени В.М. Матросова
Сибирского отделения Российской академии наук
Иркутск, Россия**

**Центр научных исследований и высшего образования
(CICESE Research Center)
Энсенада, Мексика**



**Материалы VI Международного семинара
по информационным, вычислительным и управляющим системам
для распределенных сред
(ICCS-DE 2024)**

1 – 5 июля 2024 г., Иркутск, Россия

**Иркутск: Издательство ИДСТУ СО РАН
2024**

УДК: 004.7/.9

ББК: 32.96 + 32.97

Редакционная коллегия:

Бычков И.В., Черных А.Н., Феоктистов А.Г.

Материалы VI Международного семинара по информационным, вычислительным и управляющим системам для распределенные сред (ICCS-DE 2024), 1-5 июля, 2024, Иркутск, Россия. Иркутск: Изд-во ИДСТУ СО РАН, 2024. 225 с.

Proceedings of the 6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 1-5, 2024, Irkutsk, Russia. Irkutsk: ISDCT Publisher, 2024. 225 p.

ISBN 978-5-6041814-6-1

Материалы научного сборника включают избранные статьи и тезисы VI Международного семинара по информационным, вычислительным и управляющим системам для распределенные сред (ICCS-DE 2024), проведенного Институтом динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук (Иркутск, Россия) совместно с Центром научных исследований и высшего образования (CICESE Research Center, Энсенада, Мексика) 1-5 июля, 2024 г. в г. Иркутск, Россия. Избранные научные работы посвящены развитию теории и практики разработки информационных, вычислительных и управляющих систем, относящихся к различным предметным областям информационной, технической, экономической и других сфер человеческой деятельности. Особое внимание в них уделяется системам, функционирующим в различных распределенных средах.

Официальный сайт семинара ICCS-DE 2024: <https://iccs-de.icc.ru/>

ISBN 978-5-6041814-6-1



ОРГАНИЗАТОРЫ

Международный программный комитет семинара

Председатель программного комитета

Бычков И.В., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, г. Иркутск, Россия

Сопредседатель программного комитета

Andrei Tchernykh, CICESE Research Center, Ensenada, Mexico

Члены программного комитета

Jorge Mario Cortés-Mendoza, Polytechnic University of Amozoc (UPAM), San Andrés las Vegas, Amozoc, Puebla, Mexico

Dalibor Dobrilovic, University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia

Xiwang Guo, Northeastern University, Shenyang, China

Ljubica Kazi, University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia

Zeljko Stojanov, University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia

Бабенко М.Г., Северо-Кавказский федеральный университет, Ставрополь, Россия

Еделев А.В., Институт систем энергетики им. Л.А. Мелентьева Сибирского отделения Российской академии наук, Иркутск, Россия

Ковтуненко А.С., Уфимский Государственный Авиационный Технический Университет, Уфа, Россия

Костромин Р.О., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Легалов А.И., Национальный исследовательский университет "Высшая школа экономики", Москва, Россия

Смагин С.И., Вычислительный Центр Дальневосточного отделения Российской академии наук, Хабаровск, Россия

Федоров Р.Е., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Феоктистов А.Г., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Черкашин Е.А., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Шигаров А.О., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Шокин Ю.И., Федеральный исследовательский центр информационных и вычислительных технологий, Новосибирск, Россия

Юрин А.Ю., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Организационный комитет семинара

Председатель организационного комитета

Феоктистов А.Г., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, г. Иркутск, Россия.

Заместитель председателя организационного комитета

Жарков М.Л., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Члены организационного комитета

Бабанин Д.Г., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Башарина О.Ю., Уральский государственный экономический университет, г. Екатеринбург, Россия

Велякина В.Б., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Игнатьев И.Н., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Кононенко Г.Б., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Костромин Р.О., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Кумачев А.А., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Маджара Т.И., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Максимкин Н.Н., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Мелкозеров С.В., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Попова А.К., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Столбова Г.Н., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Федоров Р.К., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Фереферов Е.С., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

Черкашин Е.А., Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, Иркутск, Россия

СОДЕРЖАНИЕ

ПЛЕНАРНАЯ СЕССИЯ

I. Bychkov, E. Fereferov, R. Fedorov. Digital Monitoring of the Environmental Situation of the Baikal Natural Territory.....	8
Z. Stojanov. Knowledge Management within Onboarding Practice in Small Software Companies	14
М. С. Хайретдинов, Г. М. Шиманская. Обратные задачи геоакустики в мониторинговых системах	22

СЕКЦИОННЫЕ ДОКЛАДЫ

A. Banshchikov. Symbolic Computation in the Analysis of the Dynamics of an Orbital Gyrostat.....	30
E. Cherkashin, N. Nagul, Y. Wang, I. Bychkov. Hierarchical Control Development Environment: Requirements and Cases	37
D. Dobrilovic, R. Dragović. Usage of Threat Modelling Tools for Implementation of ICT Security Standards in IoT Environments.....	45
A. Hmelnov. Services for Web-Publishing and Extraction of Fragments of Big Raster Data-sets.....	53
A. Shigarov, V. Paramonov. Comparison of Table Type Taxonomies	65
I. Vecstejn, Z. Stojanov, M. Kavalic, V. Gluvakov, M. Gaborov. Literature Review – Ontology Frameworks for Knowledge Management in Software Engineering	73
О. Башарина, А. Еделев, Т. Фер. Методика построения аналитической модели потребления природного газа	81
Н. Береснева, А. Еделев. Методика выбора компенсирующих мероприятий в исследованиях живучести топливно-энергетического комплекса страны.....	86
Н. Допиро. Разработка генератора платформенно-зависимой модели для системы проектирования виртуальных помощников	91
Н. Дородных, А. Юрин. Система визуального моделирования знаний Knowledge Modeling System	97
А. Еделев, Г. Данилов, О. Башарина. Разработка модели системы газоснабжения на основе открытых данных для исследования живучести	103
А. Еделев, Д. Карамов, D. V. Vinh, N. T. T. Huong. Синтез микросетей для энергоснабжения сельскохозяйственных ферм Вьетнама	111
О. Еделева, А. Еделев. Структурно-параметрическая оптимизация энергоисточников в задачах развития систем теплоснабжения	118
Д. Карамов, Е. Юмашев. Математическое моделирование солнечной радиации с учетом динамики средних и нижних слоях атмосферы	126
Р. Костромин. Подход к развертыванию программного обеспечения вычислительного кластера по облачной модели.....	135
А. Михеев, Е. Осипчук, Н. Абасов, В. Никитин. Оптимальное планирование долгосрочных режимов работы каскадов ГЭС в целях максимизации дохода генерирующей компании	141
А. Нестеров. Актуальные вопросы оценки аспектов информационной безопасности в модели интероперабельности организационно-технических систем	148

О. Николайчук, А. Павлов. Сервис автоматизации создания интеллектуальных систем на платформе Яндекс.Облако	153
Р. Федоров. Метод композиции сервисов обработки пространственных данных	160
А. Феоктистов, М. Воскобойников, А. Еделев. Методы и средства разработки и применения испытательного стенда сервис-ориентированных приложений	169
А. Хрыпченко. Разработка системы управления сбором, хранением и контролем Big Data в задачах геоакустики	174
Е. Черкашин, Н. Свердлова, Е. Марьясова. Синтаксический анализатор русского языка на основе данных морфологии: функциональный подход	182

СПЕЦИАЛЬНАЯ СЕКЦИЯ ДЛЯ МОЛОДЫХ УЧЕНЫХ

K. V. Tobola, N. O. Dorodnykh. Semantic Column Annotation for Russian Tables Using Multilingual BERT Language Model	189
В. Андрющенко. Сегментация изображений документов при помощи глубоких нейронных сетей	193
Д. Карамов, Е. Юмашев. Методы и средства реализации моделей алгоритмов анализа актинометрических показателей	195
Д. Копылов, А. Михайлов. Классификация блоков изображения на основе графового представления выделенных слов с использованием нейронных сетей	201
О. Копылова. Методы повышения точности пассивной локации импульсных источников геоакустических колебаний	205
А. А. Кузнецов, А. А. Михайлов. Защита от атак с использованием омоглифов на модели NLP	210
А. Линявский. Сравнение моделей Игрового Искусственного Интеллекта	213
М. А. Чекан. Среда программирования киберфизических систем в парадигме машин состояний	217
Д. Яковлев, М. Воскобойников, Р. Костромин. Средство мониторинга узлов системы управления сервис-ориентированными научными приложениями	220

Digital Monitoring of the Environmental Situation of the Baikal Natural Territory

Igor V. Bychkov¹, Evgeniy S. Fereferov¹ and Roman K. Fedorov¹

¹ *Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences, 134 Lermontov str, Irkutsk, 664033, Russia*

Abstract

The paper presents the results of research obtained within the framework of a large scientific project "Fundamentals, methods and technologies of digital monitoring and forecasting of the ecological situation of the Baikal natural territory". The project is aimed at creating new methods and technologies for integrated environmental monitoring and forecasting based on digital platforms that collect, store, process, analyze large arrays of heterogeneous spatial and temporal data, as well as a set of mathematical and information models, services and machine learning methods and their validation for the Baikal Natural Territory.

Keywords

Digital monitoring, forecasting, digital platforms, web-services, machine learning

1. Introduction

The Baikal Natural Territory (BNT) is a complex natural complex, where Lake Baikal, recognised by UNESCO as a World Natural Heritage Site, is located. The ecological situation in the BNT is not favourable. Large cities and industrial centres located in the southern part of the BNT are sources of air, water and soil pollution, which in turn affects the health of the population. In addition to environmental problems of anthropogenic nature, there are dangerous geodynamic processes in the BNT (seismicity, landslides, rockslides, mudflows, etc.), as well as problems with the state of boreal forests (including epidemiological damage, spread of pests, natural fires, climate change, etc.). The development of means and methods of comprehensive geo-ecological control is an important task for the conservation of Lake Baikal ecosystem and the whole BNT.

A consortium of 16 institutes of the Siberian Branch of the Russian Academy of Sciences was formed to implement the project aimed at developing new methods and technologies for integrated environmental monitoring and forecasting based on digital platforms. Consortium members have extensive practical experience in studying changes in the hydrosphere, lithosphere, biosphere, atmosphere, human health under the influence of natural and anthropogenic factors, as well as in applying mathematical methods and information-computing technologies to solve complex scientific and applied problems.

In 2024, research continued in the direction of developing a set of methods, mathematical and information models for analysis and forecasting of the Baikal natural territory development on the basis of a digital platform, which will significantly accelerate scientific research due to intellectualization, interoperability of services and their compositions.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: bychkov@icc.ru (Bychkov I.V.); fereferov@icc.ru (Fereferov E.S.); fedorov@icc.ru (Fedorov R.K.)

ORCID: 0000-0002-1765-0769 (Bychkov I.V.); 0000-0002-7316-444X (Fereferov E.S.); 0000-0002-2944-7522 (Fedorov R.K.)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.01

2. Digital platform for environmental research

To implement digital environmental monitoring of the Baikal natural territory, the ISDCT SB RAS has developed a digital platform [1-3] (Fig.1), which unites researchers from scientific institutes, employees of departmental organizations and provides them with support for integrated analysis of monitoring data, forecasting and development of recommendations for the conservation of Lake Baikal and adjacent territories. Modern approaches to the creation of digital platforms [4, 5] are usually based on the use of service-oriented architecture (SOA) [6]. This architecture allows to implement a unified information environment supporting algorithmic mutually beneficial relationships of a significant number of independent participants of any process (e.g., environmental monitoring). Within SOA, each participant provides the results of its activity in the form of services, which allows instant access to data and processing methods without the need for coordination. SOA significantly speeds up research. To unify services in the digital platform, the Web Processing Service (WPS) standard [7], which regulates the service call protocol and metadata exchange, is used by the Open Geospatial Consortium [7].

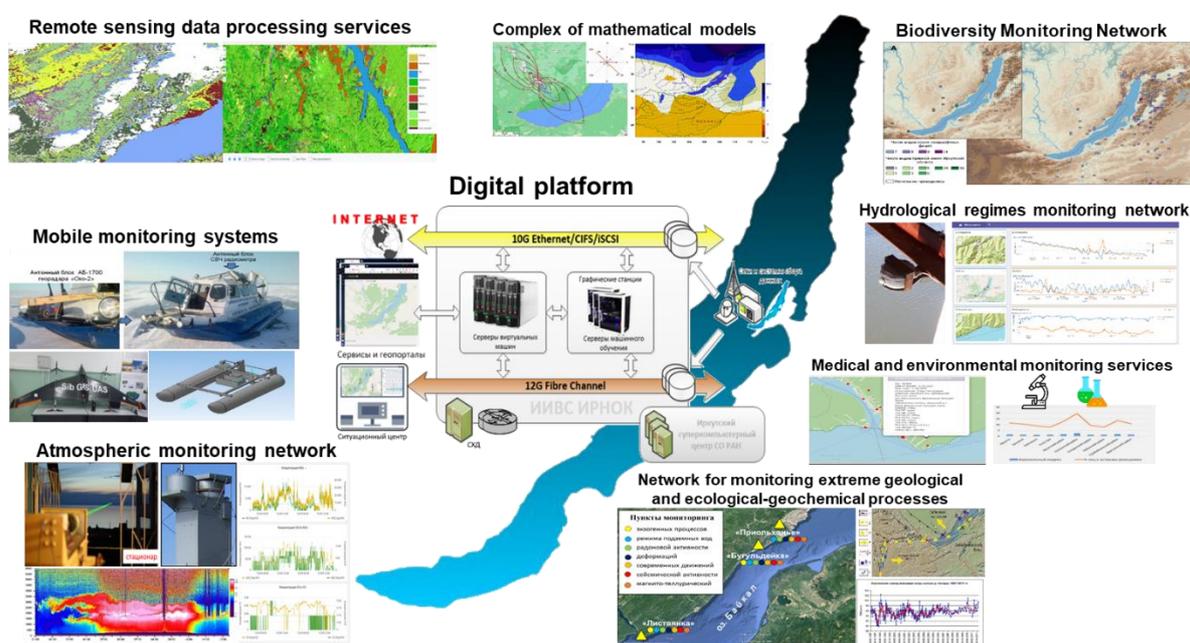


Figure 1: Digital platform for environmental research

The digital platform developed under the project acts as an integrating framework for digital environmental monitoring, is an environment for accumulation, exchange and management of data in a structured form, as well as a system for calling business functions with services of participants of the digital ecosystem (monitoring participants) connected to it through technological interfaces. The digital platform of environmental monitoring consists of three basic components: instrumental, infrastructural and application digital platforms. The Instrumental Digital Platform (IDP) provides access to the participants of the digital ecosystem of monitoring to the development and debugging of applied information and software-hardware, providing them with tool services for processing spatial and temporal data and their interfaces. The IDP includes tools for creating three main types of components embedded in a common environment:

1. development of WPS services and their compositions designed for software data processing, realizing software interoperability standards;
2. creation of data exchange services that provide and make available thematic data collection;
3. deployment of geoportals providing access to thematic data and services.

The IDP contains tools for creating generic geoportals and services in a cloud data center environment. Creation is based on pre-configured virtual machine templates. For WPS services there

are templates with Zoo project [8], 52°North Web Processing Service systems that simplify realization of WPS standard.

The Infrastructure Digital Platform (InfDP) provides the creation of applied hardware and software monitoring tools, thematic WPS services of data processing and distributed data storage based on a high-performance information and analytical environment, service-oriented and end-to-end technologies. The InfDP contains means of data delivery, processing, storage, as well as InfDP deployment infrastructure, services (software solutions based on PPI), means of development, debugging and integration of services. In order to solve the project tasks, a computing infrastructure is deployed on the basis of the IICN IREC [9], which provides all project participants with the ability to store large volumes of multiformat spatial and temporal data, high-speed access to them and the ability to perform high-loaded calculations. The infrastructure consists of a 10G-Ethernet-based high-speed data transmission network connected storage system based on Infortrend EonStor GS 1000 Gen2 (736 TB of capacity), high-performance storage system based on HPE MSA 2060 (27 TB of capacity), high-speed storage area network (SAN) operating on the basis of Fibre Channel 12G protocol, computing cluster of 15 servers with NVidia RTX- 3080, 3090 GPUs (for machine learning). On the basis of the main infrastructure components, using VMWare vSphere 7 virtualization system, a cloud platform was formed, consolidating network and computing resources with resources of data storage systems (network) and flexibly providing them for all systems and services of the project.

The applied digital platform (ADP) operates processed data at the level of a separate group or type of monitoring as a whole, as well as supports the algorithmic exchange of services between independent participants of the digital ecosystem using a unified information and analytical environment and technological infrastructure. The ADP consists of a digital monitoring geoportal, data provision services and thematic WPS services.

3. Monitoring technologies based on Earth remote sensing data

The project pays special attention to monitoring methods based on Earth remote sensing (ERS) data processing. A considerable amount of information has been collected: Sentinel-2 satellite images of the Irkutsk Region and the Republic of Buryatia for three years occupy more than 57 terabytes (21 thousand images). To organize a convenient search for space images and quick access to them, a remote sensing data catalog was created, a possibility to display images on the map and support for acquisition of space images of the Earth's surface from various sources was implemented. The images are stored in IICN IREC storage, their registration is performed by the developed data acquisition components. Image search with regard to sensor, cloud cover, position and date is implemented.

To organize environmental monitoring in large areas, a technology and Web-service (Fig.2) for classification of multispectral space images Sentinel-2 has been developed [10]. Classification of space images is carried out using the ResNet50 neural network. The technology allows to define a set of classes (surface types), prepare a training sample, verify and classify a large number of images. The training sample is generated using the Web interface based on Sentinel-2 satellite images and field surveys. As input data, the neural network accepts 13 Sentinel-2 channels and a channel built on the basis of local binary patterns to account for textural characteristics. Characteristic feature classes are identified for the BPT: Bogs, Bare Rock, Grassland, Clearings, Transitional Forest/shrubs, Coniferous Forest, Riparian Forest, Deciduous Forest, Mixed Forest, Clouds, Cropland, Residential Area, and Water. 134 space images of BNT were marked, the number of polygonal objects - 2246 pieces, the area of more than 5900 square kilometers. The average classification accuracy amounted to 95%. The developed Web-service allows to increase the efficiency of monitoring of large territories. Classification technology can be used to solve actual problems of large natural areas, in particular, to analyze changes in the forest fund (logging, fires, restoration dynamics), assess the impact of climate change on the landscape, analyze the dynamics of development, inventory of agricultural land, etc.



Figure 2: Example of the result of space images classification

4. Web services for environmental monitoring

The project has created a number of services that solve demanded tasks for the environmental monitoring of BPT. All services are implemented using the components of the instrumental digital platform and support the developed interfaces and data exchange services, which allows simplifying the integration of services and solving complex tasks. The developed services are embedded in the CP and can be applied by users for different tasks. Transaction costs, in this case time costs, are reduced due to the allocation of computing resources that are in constant access, reduction of human participation, standardization of the protocol of service invocation, etc. The created services realize modern methods of data processing and provide unique data of ecological monitoring of BNT. The following thematic services were developed to support digital environmental monitoring of the BNT and Lake Baikal:

- Monitoring service based on remote sensing data processing;
- Biodiversity monitoring service;
- Service for monitoring of hazardous geological processes;
- Atmospheric pollution monitoring service based on data from reference stations;
- Atmospheric pollution monitoring service based on Sentinel-5P satellite data (Fig.3);
- Thunderstorm activity monitoring service;
- Service for monitoring pollution of aquatic ecosystems;
- Service of forest fire risk analysis and forecasting;
- Service for registration of bites by ixodal ticks;
- Service for assessing the impact of forest fires on respiratory organs;
- Service of identification and assessment of changes in the state of vegetation cover;
- Service "Gas and aerosol impurities".
- Service "Generalized data of automatic atmospheric control stations"
- Service "Concentration of substances affecting respiratory organs"; Service "Concentration of substances affecting respiratory organs"
- Service "Atmospheric pollution index";
- Service "Catalog of earthquakes";
- Service "Fault Field";
- Service "Forest-pathological monitoring";
- Service "The burning of the forest";
- Service "Felling (forest)";
- Service "Phthalates content in water of Selenga River delta channels";

- Service "Level of pH (degree of acidity) in water by seasons";
- Service "Water turbidity by seasons of the year";
- Service "Dissolved oxygen content in water by seasons of the year";
- Service "Content of heavy metals in bottom sediments by seasons of the year";
- Service "Content of heavy metals in water by seasons of the year".

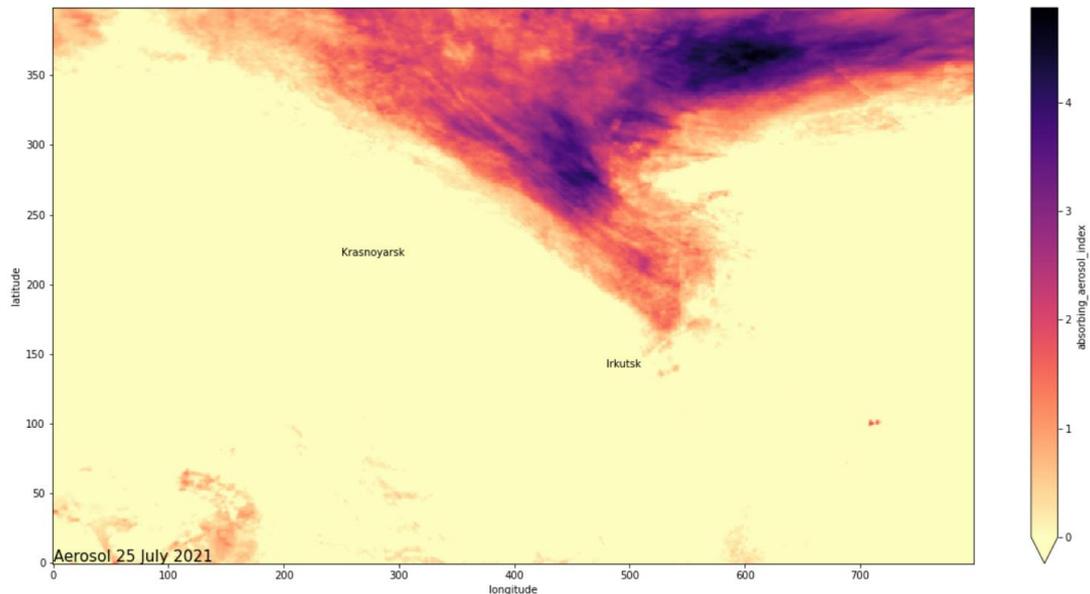


Figure 3: Example of the result of space images classification

5. Conclusion

The results obtained within the framework of the project are primarily of fundamental nature and open new opportunities in the field of development of digital technologies to support large-scale studies in large areas, processing of remote sensing data, mathematical modeling of environmental processes, monitoring and forecasting of natural and anthropogenic emissions in the atmosphere, monitoring and forecasting of hydrological regimes of water bodies, monitoring of dangerous geological processes, monitoring of biodiversity, medical and scientific research, and monitoring of biodiversity. The complex of program components is actively used in practice. More than 200 data services and more than 40 data processing services have been created. 5 different geoportals oriented to different subject areas and teams were deployed. Based on the created services, service compositions were formed, combining data services, processing and publishing services. The created digital platform for environmental monitoring of BNT is an environment for accumulation, exchange and management of data in a structured form, as well as provides the call of analytical functions with the services of participants of the digital ecosystem (monitoring participants) connected to it through technological interfaces.

6. Acknowledgements

The work was supported by the Ministry of Science and Higher Education of the Russian Federation, the grant No. 075-15-2024-533 for implementation of Major scientific projects on priority areas of scientific and technological development (the project «Fundamental research of the Baikal natural territory based on a system of interconnected basic methods, models, neural networks and a digital platform for environmental monitoring of the environment»).

7. References

- [1] Bychkov I.V., Fedorov R.K., Fereferov E.S. Instrumental components of the digital platform for environmental monitoring of the Baikal natural territory // *Computational Technologies*. 2023. Vol. 28, №6. P. 95-107. DOI: 10.25743/ICT.2023.28.6.009.
- [2] Bychkov I.V., Fereferov E.S. Digital Technologies for Monitoring and Forecasting the Environmental Situation in Siberia // *Herald of the Russian Academy of Sciences*, 2022, Vol. 92, No. 2, pp. 133–140. doi: 10.1134/S101933162202006X.
- [3] Бычков И.В., Гладкочуб Д.П., Ружников Г.М. и др. *Фундаментальные основы, методы и технологии цифрового мониторинга и прогнозирования экологической обстановки Байкальской природной территории*. 2022. Новосибирск: СО РАН, 2022. 345 с.
- [4] De Reuver M, Sørensen C and Basole R C 2018 The digital platform: A research agenda *J. Inf. Technol.* 33 124-35. doi: 10.1057/s41265-016-0033-3
- [5] Osipov Yu M, Yudina T N and Geliskhanov I Z 2018 Digital platform as an institution of the era of technological breakthrough *Economic Strategies* 5(155) 22-9
- [6] Bih, J. Service oriented architecture (SOA) a new paradigm to implement dynamic e-business solutions // *Ubiquity*. 2006. Vol. 4. P. 1–17.
- [7] WPS - Standardized web-based geo-processing. URL: <https://52north.org/software/software-projects/wps/>
- [8] ZOO-Project – Open WPS platform. URL: <http://www.zoo-project.org/>
- [9] The Integrated information and computing network of the Irkutsk scientific and educational complex. URL: <http://net.icc.ru/>
- [10] Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Popova A.K., Avramenko Y.V. Classification of Sentinel-2 satellite images of the Baikal Natural Territory // *Computer Optics*. 2022. Vol. 46, №1. pp. 90-96. doi: 10.18287/2412-6179-CO-1216.

Knowledge Management within Onboarding Practice in Small Software Companies

Zeljko Stojanov¹

¹University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Djure Djakovica BB, Zrenjanin, Serbia

Abstract

The software industry is one of the most dynamic branches of the economy in which knowledge plays an important role in achieving success and surviving in the market. Given the openness of the market, globalization, and labour turnover in the software industry, local small software companies are increasingly challenged to attract, onboard, and retain software engineers. This paper aims to present the authors experiences in inquiring onboarding practice in a local small software company, with focus on the role of knowledge management in onboarding process. The study is based on the field work in the company, with the use of qualitative research methods, ensuring that findings are grounded in the real practice. Based on the study findings, recommendations for the practice in small software companies and researchers are discussed.

Keywords

software industry, small software companies, onboarding, knowledge management

1. Introduction

In today's dynamic world, and especially a dynamic economy, with evident movement towards remote work via the Internet, software professionals can easily change jobs and move to new companies. Software companies that hire new employees should introduce them to the company's organization, processes, and organizational culture, which is commonly implemented through the *onboarding* process [1]. Selecting and onboarding the most suitable people is challenging for each software company because it influences its overall business performances [2]. In general management literature, onboarding refers to *organizational socialization* aimed at making new hires effective members of an organization by learning necessary knowledge, skills, and behaviours [3]. Anakwe and Greenhaus [4] suggested considering the following issues during organizational socialization: what information should be transmitted to new hires, which tactics to use for information transmission, how to evaluate information transmission to new hires, and what the relationships between organizational socialization and adopted behaviours and attitudes exist.

The knowledge-based perspective of companies assumes that knowledge exists in them at any time and can be used for creating new knowledge and in business actions. This promotes knowledge management (KM) as set of activities that enable identifying and leveraging the

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ zeljko.stojanov@uns.ac.rs (Z. Stojanov)

ORCID 0000-0001-6930-5337 (Z. Stojanov)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (iccs-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.02

collective knowledge in an organization for increasing their innovativeness and responsiveness in business [5]. Effective KM assumes that companies possess and improve learning capability, usually based on analyzing their successful and unsuccessful activities, leading to creating and acquiring new knowledge that may improve their practice [6].

Software industry has been recently recognized as a part of the economy with dominant influence of knowledge workers, pointing out the importance of knowledge management (KM) for both their daily practice and overall business [7]. For software organizations knowledge residing in the organization becomes a strategic asset for ensuring quality of products and services, as well as competitiveness at the market. In this course of thinking, the main challenge for software companies is to effectively involve new people in their organization, as well as to retain experienced workers who possess knowledge on all aspects of knowledge relevant for company practice (organizational, technical, technological) [8, 9].

The complexity of onboarding process in software companies is based on the fact that new hires should acquire different types of knowledge, such as organisational knowledge, project knowledge, product and domain knowledge, and knowledge of the technical environment [10]. New hires, especially those that recently graduated are usually not well prepared for real projects, which emphasises the importance of learning during the onboarding process [11]. Since onboarding period may be stressful for new employees, they should be encouraged to express themselves during the work, and to adjust to organizational identity, while preserving their own personal identity [12]. It is also important to recognize that onboarding process is an opportunity for mutual learning between the organization and new hires [13].

Small software companies make up the majority in many countries around the world, with the significant share in their economies [14]. However, small software companies suffer from the lack of resources, primarily financial, for investing in research and development to improve practice, products and services [15]. Therefore, small software companies are forced to improve their competitiveness through innovation, which is primarily based on the knowledge possessed by employees [16]. This indicates the need for small companies to systematize the knowledge that exists in them through organizational learning [17].

Based on the previous considerations, KM in onboarding practice is especially challenging for small software companies that strive to acquire and retain knowledgeable workers. This paper aims to present experiences in inquiring KM in onboarding practice in a small software company, and based on that to propose recommendations for practitioners in industry and researchers.

2. Experience in a Local Small Software Company

The presented experience and the findings are part of a larger project designed and implemented by the research team composed of people from the *Information Technology Department* and *Engineering Management Department* from Technical faculty "Mihajlo Pupin" Zrenjanin, Serbia. The project objective is the inquiry into onboarding and knowledge management practices in small software companies in Serbia, with a special focus on assisting indigenous software companies in practice improvement and better positioning in the market.

In this section will be presented results of joint practical and research work of the author and

the company manager of a local small software company. The research objective is to identify knowledge management practice relevant for onboarding related processes in the company.

The study uses qualitative research methods with flexible research design [18] suitable for inquiring human related issues in the industrial practice [19, 20, 21] or systematizing knowledge about the practice [22]. Reflexive thematic analysis [23] for analysing data and constructing the findings is used.

2.1. Context

The presented research relates to the practice in a local small company that provides software engineering services and IT infrastructure services according to the turnkey principle. The company has 8 employees and has operated for 20 years in Serbia and the region. Software engineering services include the development and maintenance of specific software solutions, mostly based on cloud and mobile technologies. The main domains of applications are inventory management and monitoring of fixed assets, control of the implementation of standards in the company, including the qualifications of employees for optimal engagement in the workplace and the organizational structure with the systematization of workplaces.

2.2. Knowledge Management in Onboarding Practice

This section will present findings related to knowledge management practices within onboarding in a selected small software company. The findings are constructed through an inductive approach based on reflexive thematic analysis and are grounded in interviews with the company manager and available documents. The findings are developed as a thematic framework that includes themes and relationships between them, representing identified issues in the observed practice in the company.

Through thematic analysis a thematic framework with several main themes related to onboarding practice was developed. The whole thematic framework contains the following themes [24]: *Onboarding process*, *Objectives*, *Prerequisites*, *Outputs*, *Shortcomings*, *People*, *Tools and methods*, *Work position*, and *Knowledge management*. Each of the main themes encompasses several sub-themes related to a specific aspect or segment of the theme.

The segment of the findings related to relationships between onboarding process and knowledge management is presented in Figure 1. The direct relationship between two main themes (*Onboarding process*, *Knowledge management*) is shown with a thick solid line, while indirect relationships through other themes (*Work position*, *Tools and methods*) are shown with thin dashed lines.

Onboarding process was identified as the main and central theme in the thematic framework, relating to the process all candidates for working engagement should pass. The steps in the onboarding process are: getting to know the overall organization and culture of the company, introduction to a team in which a candidate will work, introduction to a technical platform, and introduction to projects in which a candidate will be involved.

Knowledge management is the next most important theme related to issues of capturing, organizing, and reusing various types of knowledge that exist in the company. The main characteristic of knowledge management is that it is managed selectively, which means that

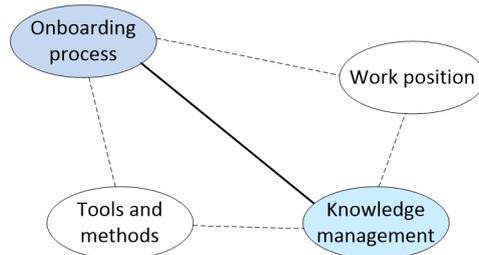


Figure 1: Relationships between onboarding practice and knowledge management in the selected small software company

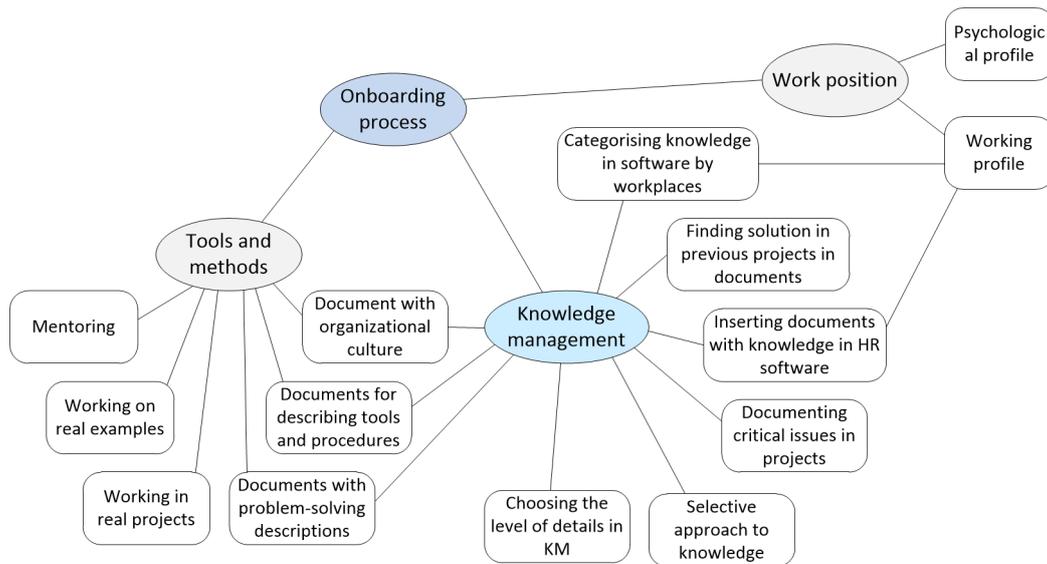


Figure 2: Detailed thematic framework with themes and subthemes related to the relationships between onboarding practice and knowledge management

only specific and the most critical knowledge from past projects was selected, collected, and prepared for current and future projects.

Tools and methods is another theme related to practical issues in onboarding process realization. It includes the use of various types of documents in onboarding (documents with organizational culture, documents with technical issues and problems), mentoring candidates, and assigning simple tasks from real projects to candidates.

Work position is the theme related to tailoring the onboarding process to specific work positions in the company, which includes determining the working profile (technical and technological issues) and psychological profile of candidates.

The developed thematic framework contains the main themes represented in elliptical shapes and subthemes represented in rounded rectangles, as it is presented in Figure 2.

The relationship between theme *Knowledge management* and theme *Work position* is realized through the connections of subtheme *Working profile* associated to *Work position* and subthemes

Categorising knowledge in software by workplaces and Inserting documents with knowledge in HR software associated to *Knowledge management*. These themes and relationships are grounded in the company practice since the company maintains a working profile with associated knowledge for each work position in the HR software. During the onboarding process, the new employees are directed to use the software to access knowledge for their positions. As the practice evolves, documents with knowledge change and adapt to new working conditions, leading to updates of the knowledge base in the software.

The relationship between theme *Knowledge management* and theme *Tools and methods* is realized through identification of three subthemes that are associated to both main themes. These subthemes point out the use of various types of documents for systematizing knowledge (*Document with organizational culture*, *Documents for describing tools and procedures*, and *Documents with problem-solving descriptions*), and at the same time these subthemes are identified as a tool within onboarding process in the company. These subthemes related to the use of documents identify the following aspects of knowledge to be managed for successful onboarding practice in the company:

- *Knowledge on organizational culture*. This aspect reflect the importance of human and organizational issues in the company, to which new employee should align during onboarding process.
- *Knowledge on tools and procedures in technical work*. Technical part of the work consumes majority of time in the company, and quick and effective introduction to technical tools (integrated development environments, platforms for sharing materials, version control systems, etc.) and working procedures is essential for onboarding and overall business performance in the company.
- *Knowledge on problem-solving issues*. Software development, as a set of knowledge based activities, requires engineers with problem-solving skills. Having systematized experience with problems in previous projects can significantly improve onboarding practice.

3. Recommendations

3.1. Recommendations for small software companies

Based on the analysis of literature in the introduction section, and findings of the study, the following recommendations for the managers and engineers from small software companies can be proposed:

- *Hiring academic researchers for practice research*. Lack of resources, employees fully devoted to daily tasks, and the lack of experience with research methods by employees is the reason for not starting projects for analysing and improving practice in small companies. The most effective way to overcome this is to engage researchers with experience and work together on a research project.
- *Use of established and well known knowledge management methods and tools*. Avoiding ad-hoc approaches through the use of proven methods and tools during the inquiry of practice will result with more reliable KM practice and its better alignment with onboarding process.

- *Aligning knowledge management with all working positions.* Knowledge should be carefully inquired and systematized for each particular working position. This should include knowledge relevant to the psychological, organizational, and technical aspects of each working position.
- *Systematizing and documenting knowledge residing in the company.* Knowledge should be collected from previous projects, activities, and tasks, and made explicitly available to all employees, especially those in the process of onboarding. This will make the onboarding process more effective, while minimizing engagement of other employees in the company during onboarding.
- *Connecting the onboarding process with real projects to enhance the acquisition of real knowledge.* Including new hires during onboarding in real projects enables them to gain knowledge on the most relevant technical and non-technical issues. This will enable the direct sharing of knowledge between employees with more experience and new hires, and foster organizational learning in the company. It also leads to an increase in the satisfaction of new employees due to their contribution in solving real problems on projects.

3.2. Recommendations for researchers

In addition to recommendations to practitioner from software industry, the following recommendations can be proposed for researchers:

- *Joint preparation and implementation of the research with company employees.* The best possible overview of the practice and the most relevant research objective can be achieved through joint work with company employees in all phases of the research. In addition, this will ensure more reliable findings.
- *Research based on active field work.* Active fieldwork presupposes the presence of the researchers in the real world environment using proven fieldwork methods such as interviews, observation of practice and note-taking, access to documentation and electronic data, and meetings. This enables a more comprehensive insight into practice, and increases the confidentiality of findings by triangulating data sources and applied research methods.
- *Involvement of company employees in all research activities.* As already emphasized in the literature, employees are the main source of knowledge about organizational processes, organizational culture and technical aspects of practice, so it is necessary to actively include them as sources of data, as participants in data analysis, but also later when verifying research findings.

4. Conclusions

This paper points out the importance of KM and onboarding practice in small software companies, which needs to be creative and responsive to survive at the competing market. In addition to presenting well known issues reported in the relevant scientific literature, the paper presents the experience in inquiring onboarding practice in the local small software company. Research

findings, constructed through qualitative inquiry of the company's practice, indicate multiple relationships between KM and onboarding in the selected company. Based on the study findings, recommendations for small software companies, and researchers are proposed.

References

- [1] G. G. Sharma, K.-J. Stol, Exploring onboarding success, organizational fit, and turnover intention of software professionals, *Journal of Systems and Software* 159 (2020) 110442. doi:10.1016/j.jss.2019.110442.
- [2] T. Hall, S. Beecham, J. Verner, D. Wilson, The impact of staff turnover on software projects: the importance of understanding what makes software practitioners tick, in: *Proceedings of the 2008 ACM SIGMIS CPR Conference on Computer Personnel Doctoral Consortium and Research*, Charlottesville, VA, USA, 2008, pp. 30–39. doi:10.1145/1355238.1355245.
- [3] T. N. Bauer, B. Erdogan, Organizational socialization: The effective onboarding of new employees, in: S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology*, Vol. 3. Maintaining, expanding, and contracting the organization, American Psychological Association, 2011, pp. 51–64. doi:10.1037/12171-002.
- [4] U. P. Anakwe, J. H. Greenhaus, Effective socialization of employees: Socialization content perspective, *Journal of Managerial Issues* 11 (1999) 315–329.
- [5] M. Alavi, D. E. Leidner, Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues, *MIS Quarterly* 25 (2001) pp. 107–136.
- [6] P. Sok, A. O'Cass, K. M. Sok, Achieving superior {SME} performance: Overarching role of marketing, innovation, and learning capabilities, *Australasian Marketing Journal (AMJ)* 21 (2013) 161–167. doi:10.1016/j.ausmj.2013.04.001.
- [7] A. K. Valacherry, P. Pakkeerappa, Knowledge management in the software industry: Creating value through knowledge application, *Journal of Creating Value* 6 (2020) 249–270. doi:10.1177/2394964320968981.
- [8] A. M. Saks, K. L. Uggerslev, N. E. Fassina, Socialization tactics and newcomer adjustment: A meta-analytic review and test of a model, *Journal of Vocational Behavior* 70 (2007) 413–446. doi:10.1016/j.jvb.2006.12.004.
- [9] V. Brødsjø, B. Sandøy, E. Hustad, Exploring onboarding processes for it professionals: The role of knowledge management, in: *European Conference on Knowledge Management*, Lisboa, Portugal, 2023, pp. 148–158. doi:10.34190/eckm.24.1.1370.
- [10] P. Gregory, D. E. Strode, H. Sharp, L. Barroca, An onboarding model for integrating newcomers into agile project teams, *Information and Software Technology* 143 (2022) 106792. doi:10.1016/j.infsof.2021.106792.
- [11] R. Pham, S. Kiesling, L. Singer, K. Schneider, Onboarding inexperienced developers: struggles and perceptions regarding automated testing, *Software Quality Journal* 25 (2017) 1239–1268. doi:10.1007/s11219-016-9333-7.
- [12] D. M. Cable, F. Gino, B. R. Staats, Reinventing employee onboarding, *MIT Sloan Management Review* 54 (2013) 23–28.
- [13] D. Jeske, D. Olson, Onboarding new hires: recognising mutual learning opportunities, *Journal of Work-Applied Management* 14 (2022) 63–76. doi:10.1108/JWAM-04-2021-0036.

- [14] M. Tuape, Y. Ayalew, Factors affecting development process in small software companies, in: Proceedings of 2019 IEEE/ACM Symposium on Software Engineering in Africa (SEiA), Montreal, QC, Canada, 2019, pp. 16–23. doi:10.1109/SEiA.2019.00011.
- [15] H. Huang, F. Li, Innovation climate, knowledge management, and innovative work behavior in small software companies, *Social Behavior and Personality: an international journal* 49 (2021) 1–17. doi:10.2224/sbp.9780.
- [16] M. Zieba, E. Bolisani, M. Paiola, E. Scarso, Searching for innovation knowledge: insight into kibs companies, *Knowledge Management Research & Practice* 15 (2017) 282–293. doi:10.1057/s41275-017-0050-1.
- [17] Z. Stojanov, Thematic knowledge framework on human factor in software maintenance practice: A study in a micro software company, *Journal of Software Engineering & Intelligent Systems* 4 (2019) 41–57.
- [18] C. Marshall, G. B. Rossman, *Designing Qualitative Research*, 6th ed., SAGE Publications, Thousand Oaks, US, 2016.
- [19] C. B. Seaman, Qualitative methods in empirical studies of software engineering, *IEEE Transactions on Software Engineering* 25 (1999) 557–572. doi:10.1109/32.799955.
- [20] Z. Stojanov, Qualitative research on practice in small software companies, in: M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*, 3rd ed., IGI Global, Hershey, PA, USA, 2015, pp. 650–658. doi:10.4018/978-1-4666-5888-2.ch062, dOI: 10.4018/978-1-4666-5888-2.ch062.
- [21] Z. Stojanov, Adopting qualitative social science methods for inquiring software engineering industrial practice, in: Proceedings of the 5th International Workshop on Information, Computation, and Control Systems for DistributedEnvironments (ICCS-DE 2023), Irkutsk, Russia, 2023, pp. 8–18.
- [22] Z. Stojanov, J. Stojanov, D. Dobrilovic, Knowledge discovery and systematization through thematic analysis in software process assessment project, in: Proceedings of the IEEE 13th International Symposium on Intelligent Systems and Informatics, SISY 2015, Subotica, Serbia, 2015, pp. 25–30. doi:10.1109/SISY.2015.7325405.
- [23] V. Braun, V. Clarke, *Thematic Analysis: A Practical Guide*, SAGE Publications, London, UK, 2022.
- [24] Z. Stojanov, S. Bajic, M. Kavalic, S. Mitic, M. Nikolic, A qualitative study on onboarding in a small software company, in: Proceedings of the XIV International conference on Information Technology and Education Development (ITRO 2023), Zrenjanin, Serbia, 2023, pp. 221–226.

Обратные задачи геоакустики в мониторинговых системах

Хайретдинов М.С.^{1,2}, Шиманская Г.М.¹

¹ *Институт вычислительной математики и математической геофизики СО РАН, пр. акад. Лаврентьева, 6, 630090, Новосибирск, Россия*

² *Новосибирский государственный технический университет, пр. Карла Маркса 20, 630092, Новосибирск, Россия*

Аннотация

Рассматривается проблема геомониторинга окружающей среды в связи с обнаружением и определением местоположения разного класса импульсных источников, несущих угрозу для окружающей социальной инфраструктуре. В качестве таких источников могут выступать различные карьерные и промышленные взрывы, падающие на землю отработанные ступени ракет при спутниковых запусках, космические тела и др., являющиеся мощными источниками сейсмических и акустических волн.

В рамках обозначенной проблемы рассматриваются задачи, связанные с обнаружением и определением местоположения разного класса источников на основе регистрируемых от них сейсмических и акустических колебаний/. Решение задачи рассматривается как совмещенной обратной задачи по исходным данным от сети пространственно-распределенных датчиков с использованием кинематических характеристик в виде времен вступлений волн сейсмических и акустических волн. Проведён сравнительный анализ результатов численных и натурных экспериментов по решению задачи геолокации источников различными методами. Разработан соответствующий программный комплекс.

Ключевые слова

Геомониторинг, взрывоподобные источники, обратная задача, геолокация, численное моделирование, полевой эксперимент.

1. Введение

Одна из основных проблем геоэкологии окружающей среды связана с оцениванием геоэкологических рисков для окружающей социальной среды и, прежде всего для человека, порождаемых мощными источниками импульсной природы. В качестве таковых могут выступать различные карьерные, промышленные, испытательные взрывы, падающие на землю отработанные ступени ракет при спутниковых запусках, космические и воздушные тела разной природы и др., являющиеся мощными источниками сейсмических и акустических волн. В частности, обусловленность экологически важной задачи обнаружения, определения координат и утилизации падающих на Землю отработанных ступеней ракет при космических запусках связана с наличием в таких телах вредного для человека гептила. Особенно это актуально для районов восточного Казахстана и Алтая при космических запусках с полигона Байконур [1].

К мониторинговым относятся также задачи, связанные с обнаружением и определением мест падения на землю космических тел типа осколков метеоритов, а также других воздушных тел разной природы.

Рассматриваемые задачи связаны с общей проблемой геолокации и идентификации импульсных источников по исходным данным от сети пространственно-распределенных датчиков с использованием кинематических характеристик в виде времен вступлений сейсмических и акустических волн. и их динамических характеристик в виде спектральных функций, соотнесенных к волновым формам колебаний от источников.

Ряд задач в связи с рассматриваемой проблемой связаны с предупреждением разрушительных последствий взрывов для прилегающей жилой инфраструктуры в районах

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: marat@opg.sccc.ru



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.03

проведения угледобычи открытым способом. Особенно это характерно для угледобывающего района Кузбасса. В рассматриваемых ситуациях экологические риски, оцениваемые удельной плотностью акустической энергии, могут многократно возрастать (до 50 раз и более) из-за развивающегося явления пространственной фокусировки геоакустических волновых полей в заданном азимутальном направлении в результате взаимодействия с метео-факторами (направления и скорости ветра, влажности и др.) [2,3], а также характерного покрова земной поверхности [4]. Эти факторы в числе других определяют актуальность решения рассматриваемых задач.

В работе рассмотрены подходы к решению таких задач. Эффективность их по критериям точности геолокации источников подкреплена результатами численного моделирования и полевых экспериментов.

2. Постановка задачи и пути решения.

В математической постановке решение рассматриваемых задач в общем виде сводится к решению обратных задач восстановления параметров источника и определения его типа по результатам измерения времён прихода волн на сети регистрирующих датчиков. Задача оценивания неизвестных параметров источника сводится к решению нелинейной системы условных уравнений [5]:

$$\vec{t} = \vec{\eta}(X, \theta) + \vec{\varepsilon}, \quad (1)$$

где $\vec{t} = (t_1, t_2, \dots, t_N)$ – вектор времён вступлений сейсмических волн; $\vec{\eta}(X, \theta)$ – N -мерный вектор вычисляемых времён вступлений (теоретический годограф) или функция регрессии; $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_N)^T$ – вектор невязок, $\vec{\theta} = (x, y, z, v, t)^T$ – m -мерный вектор оцениваемых параметров; $X = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)$ – матрица координат датчиков (или точек излучения); N – число датчиков (или точек излучения). В качестве оцениваемых параметров выступают пространственные координаты источника – x, y, z , скоростная характеристика среды v и время в источнике t_u . В ряде случаев скорость в среде является известной. При оценивании параметров пользуются сведениями о распределении ошибок $\varepsilon_i = t_i(\vec{x}_i, \vec{\theta}) - \eta(\vec{x}_i, \vec{\theta})$. Будем в дальнейшем предполагать, что ε_i – взаимно независимые случайные величины, имеющие распределение с нулевым средним и заданными дисперсиями:

$E\varepsilon_i = 0$, $E\varepsilon_i\varepsilon_j = \sigma_i^2\delta_{ij}$, $\sigma_i = \sigma(\vec{x}_i)$, δ_{ij} – символ Кронекера, $i = 1, 2, \dots, N$. В случаях затруднений с заданием дисперсий, можно принять их равными и получить несмещённую оценку дисперсии наблюдения с единичным весом.

Задача оценивания параметров θ в (1) является частью так называемого регрессионного анализа, а ее решением служат оценки метода наименьших квадратов:

$$\vec{\theta} = \arg \min_{\vec{\theta} \in \Omega} Q(\vec{\theta}), \quad Q(\vec{\theta}) = \sum_{i=1}^N \frac{(t_i - \eta_i(\vec{\theta}, v, X))^2}{\sigma_i^2} \rightarrow \min \quad (2)$$

Для отыскания минимума функционала существуют различные методы, например, итерационный метод Гаусса-Ньютона [6], метод сингулярного разложения или обобщённого псевдообращения [7], метод Качмажа [8] и др.

Целью настоящей работы является выбор и проведение исследования метода, отвечающего требованию решения поставленной задачи в мониторинговых системах реального времени, поскольку именно такое требование часто формулируется как основное.

При решении задачи используется линейная аппроксимация функции регрессии (3) в окрестности точки $\vec{\theta}^k$:

$$J(X, \vec{\theta}^k) \Delta \vec{\theta}^k + \vec{\eta}(X, \vec{\theta}^k) - \vec{t} + \vec{\varepsilon} = 0, \quad (3)$$

куда входит матрица параметров

$$J(X, \theta) = \left(\frac{\partial \eta(\vec{x}_i, \vec{\theta})}{\partial \theta_1}, \dots, \frac{\partial \eta(\vec{x}_i, \vec{\theta})}{\partial \theta_m} \right), \quad i = \overline{1, N}. \quad (4)$$

Наличие внешних шумов, приводящее к погрешностям оценивания вектора $\vec{t} = (t_1, t_2, \dots, t_N)$,

делает задачу (1) некорректной. Один из подходов к решению для этого случая связан с применением вычислительной схемы метода Гаусса-Ньютона с регуляризацией [9]. При этом требуется подбор коэффициента регуляризации α с учетом характера шумов и выработки последующего правила останова итерационного процесса. В полевых условиях это часто затруднительно. Ниже рассмотрены другие методы решения, более приближенные к требованию анализа в темпе реального времени.

2.1. Применение метода псевдообращения

Другой подход к решению задачи (2)–(4) состоит в непосредственном решении на каждом шаге процесса системы (4), для решения которой наибольшее распространения в настоящее время получил метод псевдообращения (или обобщенного обращения), основанный на сингулярном разложении (SVD-разложении) [7]. Метод имеет широкое применение. В частности, он успешно использовался авторами при решении задачи определения положения скважинного источника по глубине с применением наземной сети распределенных датчиков [5]. При этом точность определения местоположения составила не хуже 1%. Также он был рассмотрен при решении задачи определения источника в волновом уравнении по дополнительной информации, измеренной на различных частях границы области [10] и др. [6].

В силу теоремы о сингулярном разложении [7] существуют такие ортогональные матрицы U_{mn} размера $m \times m$ и V_{mn} размера $n \times n$ и набор неотрицательных чисел $\{\rho_j\}$, позволяющие представить разложение матрицы (4) на каждом k -ом шаге итерационного процесса в произведение трех матриц

$$J(X, \bar{\theta}^k) = U_{mn}^{(k)} \Sigma_{mn}^{(k)} V_{mn}^{(k)T}, \quad (5)$$

где $V_{mn}^{(k)T} = V_{mn}^{(k)-1}$, Σ_{mn} - диагональная матрица сингулярных чисел вида

$$\Sigma_{mn} = \begin{pmatrix} \rho_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_2 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \rho_m & 0 & 0 \end{pmatrix}, \rho_{j+1} \leq \rho_j$$

Метод предусматривает также проведение так называемого сингулярного анализа, состоящего в исключении нулевых сингулярных чисел и соответствующих им столбцов матриц U и V . Итерационный процесс в этом случае имеет вид:

$$\bar{\theta}^{k+1} = \bar{\theta}^k + V_k \Sigma_k^{-1} \bar{d}^k, \quad k = 0, 1, 2, \dots, \quad (6)$$

где \bar{d}^k - вектор, состоящий из первых m компонент вектора $U_k^T \bar{y}(\bar{X}, \bar{\theta}^k)$. При реализации этого процесса попутно получается не только ковариационная матрица пространства параметров, но и ковариационная матрица пространства данных. Также вычисляются матрица разрешенности $V_k V_k^T$, близость которой к единичной матрице указывает на степень разрешенности задачи, и матрица информационной плотности $U_k U_k^T$, близость которой к единичной матрице указывает на относительную значимость отдельных наблюдений [11].

2.2. Применение адаптивного метода Качмажа.

Достоинством выше рассмотренных итерационных методов решения обратной задачи является его способность получить искомое решение и оценку ковариационной матрицы за один проход по всем уравнениям, что позволяет одновременно оценить погрешности решений. Однако они, как и многие другие методы, связанные с обращением и умножением матриц, позволяют решать системы с ограниченным числом неизвестных. Это обусловлено следующими причинами: с ростом числа уравнений растут ошибки, связанные с умножением матриц; время счета растет пропорционально n^3 ; память, необходимая для хранения ковариационных матриц, растет пропорционально n^2 . С учетом этого прямые итерационные

методы решения линейных систем с матрицами большой размерности, как например, метод Гаусса-Ньютона или метод псевдообращения, в системах реального времени использовать затруднительно из-за указанных ограничительных факторов по времени и памяти. В связи с этим возникает необходимость обращения к методу (в классе итерационных), который был бы лишен этих недостатков. Этому требованию соответствует метод Качмажа [12]. Важное достоинство метода связано с тем, что уточнение неизвестных ведется по каждому алгебраическому уравнению отдельно и, соответственно, отпадает необходимость в трудоемких матричных преобразованиях. Доказаны сходимость и регуляризующие свойства метода [8]. Ниже приводится описание процесса вычислений с использованием метода Качмажа, результаты его применения интересах решения обратной задачи в геомониторинговых системах. Предварительно рассматривается значение невязки как разность между фактическим и прогнозным временами прихода прямой волны от источника к i -му приемнику $\Delta t_i = t_i - \eta(\vec{x}_i, \vec{\theta})$. Ненулевое значение ее может быть обусловлено как погрешностью в наблюдаемом времени из-за влияния внешних шумов, так и погрешностью из-за неточности описания модели среды распространения волн. В первом случае уменьшение погрешности может быть достигнуто за счет применения цифровой фильтрации волновых форм из шумов. Во втором случае неадекватность модели удастся выявить в процессе тестирования алгоритма решения обратной задачи. Учитывая оба приведенных фактора и принимая во внимание выражение можно представить невязку по времени в следующем виде:

$$\Delta t_i = \sum_{j=1}^m J_{ij} \Delta \theta_j + \varepsilon_i, \quad (7)$$

где i – номер уравнения, j – номера неизвестных, $\Delta \theta_j$ – неточность описания прогнозных параметров модели, т.е. значений неизвестных, ε_i – ошибка в наблюдаемом времени прихода волн. Тогда определение неизвестных параметров сводится к задаче оценки $\Delta \theta$ по невязке Δt . В общем случае задача имеет множество решений. Для ее разрешения будем предполагать, что неизвестные $\Delta \theta$ случайны, так же, как и ε_i . С учетом использования метода максимального правдоподобия оценки для $\Delta \theta$ на k -ом шаге для j -го неизвестного имеют вид:

$$\Delta \theta_j^k = \frac{\Delta t_i J_{ij} (\sigma_j^2)^k}{\sigma_{t_i}^2 + \sum_{j=1}^m J_{ij}^2 (\sigma_j^2)^k}. \quad (8)$$

Здесь j – порядковый номер неизвестного; i – порядковый номер уравнения; $(\sigma_j^2)^k$ – оценка дисперсии неизвестного на k -ом шаге; $\sigma_{t_i}^2$ – дисперсии ошибки измерения входного параметра в i -ом уравнении.

После получения на очередном шаге величины $\Delta \theta$ новое значение неизвестного параметра, обеспечивающего взвешенный минимум невязки, определяется итерационным уравнением:

$$\theta_j^{(k+1)} = \theta_j^k + \Delta \theta_j^{(k+1)} = \theta_j^k + \alpha_j^{(k+1)} \Delta t_i, \quad \alpha_j^k = \frac{J_{ij} (\sigma_j^2)^k}{\sigma_{t_i}^2 + \sum_{j=1}^m J_{ij}^2 (\sigma_j^2)^k} \quad (9)$$

При $(\sigma_j^2)^k = 1$ и $\sigma_{t_i}^2 = 0$ уравнение (10) сводится к уравнению, описывающему метод Качмажа [12]:

$$\theta_j^{(k+1)} = \theta_j^k + \Delta \theta_j^{(k+1)} = \theta_j^k + \frac{J_{ij}}{\sum_{j=1}^m J_{ij}^2} \Delta t_i, \quad (10)$$

Процесс уточнения (10) повторяется по каждому из уравнений параллельно несколько раз, что определяется числом итераций.

В работе [12] показаны условия сходимости. Главное из них — равенство нулю математического ожидания погрешностей времен прихода волн в исходных годографах. Всякие систематические погрешности неизбежно приведут к смещению оценок координат источника

или параметров среды. Тем не менее, такие смещения характерны и для рассмотренных ранее методов.

2.3. Численное моделирование методов геолокации

Предлагаемый подход к решению задачи геолокации решается в совмещенном варианте, т.е. когда одновременно учитываются параметры сейсмических и акустических волн в виде их времен вступления. Для этого случая функция регрессии будет иметь вид:

$$\vec{\eta}(\vec{\theta}, \vec{v}, \vec{X}) = \vec{t}_a - \vec{t}_s, \quad (11)$$

где \vec{t}_s и \vec{t}_a – векторы времен вступлений сейсмической и акустической волн соответственно, $\vec{v} = (v_s, v_a)$ – скорости сейсмической и акустической волн. С учетом известного соотношения

[13] $R \frac{v_s - v_a}{v_s v_a} = t_a - t_s$, где R – расстояние «источник-приемник» функция регрессии формируется в виде:

$$\eta_i(\vec{\theta}, \vec{v}, \vec{x}_i) = \eta_i(\vec{x}^*, \vec{v}, \vec{x}_i) = \frac{R(\vec{x}^*, \vec{x}_i)(v_s - v_a)}{v_s v_a}, i = \overline{1, N} \quad (12)$$

Выполнено численное моделирование [14,15] для решения задачи геолокации источников с геометрией их расстановки, а также сети датчиков в виде креста, как это показано на рис.1. Целью моделирования является оценивание потенциальных погрешностей определения координат источников с расстановкой вида, соответствующей рис.1, обоими рассмотренными методами-псевдообращения в соответствии с (6) и Качмажа в соответствии с (12) с последующим сопоставлением результатов моделирования. Исходными данными являются времена вступления волн на сети датчиков и определенная из полевого эксперимента скорость распространения сейсмических волн.

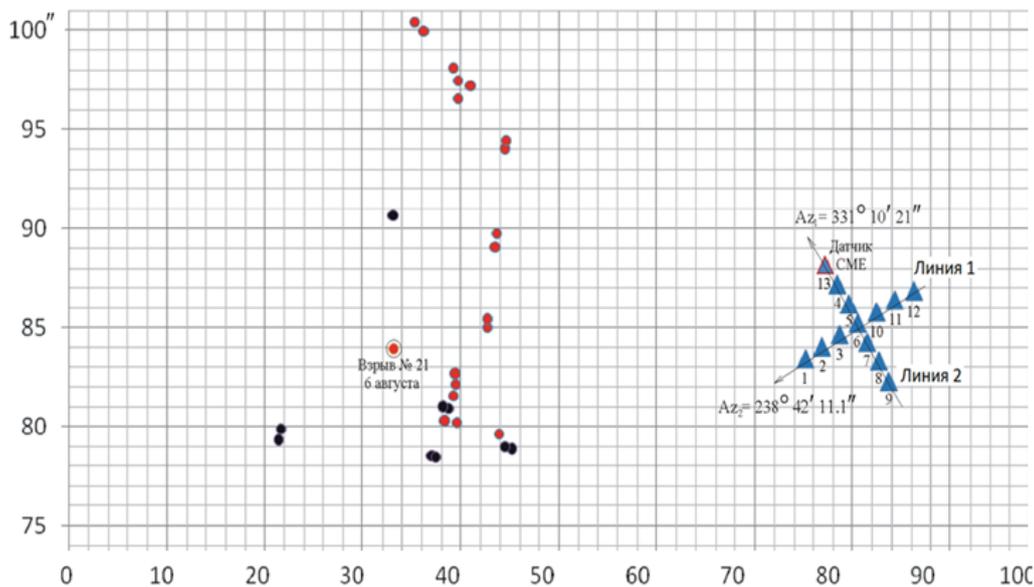


Рисунок 1: Схема взаимного расположения датчиков и импульсных источников (тестовых взрывов); треугольниками помечены датчики, а окружностями – источники.

Создана вычислительная программа в среде MatLab для реализации расчетов согласно (7). Расчёты по описываемой программе состоят в следующем: вводятся известные координаты 21 точки взрывов, затем по известной (определённой из эксперимента) скорости сейсмической волны рассчитываются времена её пробега от точек взрыва до сейсмодатчиков, Далее на эти времена «набрасываются» случайные погрешности, распределенные по гауссовому закону

распределения в пределах от -5% до +5% от времени пробега и уже эти времена задаются программе в качестве времен вступлений сейсмических волн от взрывов на сейсмодатчиках. Программа рассчитывает для каждого взрыва выходные данные: расстояние до источника – в метрах, направление на источник – в градусах, невязка (погрешность определения координат расстояния) – в метрах. Программа работает в прямоугольной системе координат.

На основе разработанной технологии моделирования геолокации была набрана статистика измерений для обоих методов измерений, представленная в таблице 1. Из таблицы 1 видно, что метод псевдообращения и адаптивный метод в результате дают примерно одинаковую погрешность. По количеству итераций эти методы также эквивалентны.

2.4. Тестирование методов на реальных данных

Для сопоставления с результатами численного моделирования выполнены расчеты в интересах геолокации обоими рассматриваемыми методами по данным реальных полевых экспериментов, выполненных по схеме рис. 1.

Задача формулируется в такой постановке: дана схема взаимного расположения датчиков и тестовых взрывов. Координаты датчиков, а также скорости сейсмической и акустической волн и их времена прихода на датчики известны. Необходимо вычислить положение каждого из источников.

Таблица 1

Расстояние, м	Азимут, °	Метод псевдообращения			Адаптивный метод		
		Расстояние, м	Азимут, °	Невязка, м	Расстояние, м	Азимут, °	Невязка, м
597,46	251,841	600,38	252,22	4,13	601,01	251,12	3,61
606,79	251,423	605,4	251,96	2,54	607,07	252,01	2,98
680,47	258,706	679,7	257,96	3,05	681,56	258,07	3,5
777,28	282,505	780,44	281,34	5,04	778,79	283,38	3,65
611,71	253,488	613,58	255,01	3,58	610,85	254,76	3,51
671,29	260,159	674,51	259,29	4,84	673,82	261,57	6,71
664,36	261,604	664,01	260,87	3,56	664,77	262,3	3,69
663,29	263,089	661,20	264,12	5,13	661,45	262,76	4,48
605,97	269,303	608,15	269,66	5,64	607,56	268,14	5,87
606,55	270,475	602,04	271,54	6,21	603,51	269,98	5,84
605,12	281,203	603,26	280,31	5,67	603,69	279,20	5,01
606,77	283,251	610,00	285,07	7,04	609,23	284,15	6,52

Данный эксперимент отличается от модельного наличием шумов с повышенным фоном от воздействия проходящего транспорта, что несомненно ухудшает точность геолокации. В таблице 2 в качестве примера приведены результаты, полученные обоими рассматриваемыми методами для части экспериментов.

Таблица 2: Сравнение результатов работы методов на реальных данных

Расстояние		597,46	606,79	680,47	777,28	611,71	671,29	664,36
Азимут		251,84	251,42	258,70	282,50	253,48	260,1	261,6
Метод псевдообращения	Расстояние, м	545,49	650,81	712,32	812,12	649,41	707,3	699,3
	Направление, °	254,21	255,5	262,3	286,87	257,02	264,1	258,3
	Количество итераций	32	38	40	35	35	31	43
	Невязка, м	42,87	51,32	34,67	42,23	37,56	44,51	35,94

Адаптивный метод	Расстояние, м	546,72	581,93	652,15	740,33	650,32	710,0	645,4
	Направление, °	248,77	248,3	249,87	278,82	258,58	265,2	269,3
	Количество итераций	29	25	29	23	19	31	30
	Невязка, м	42,59	49,13	30,22	40,4	29,36	43,35	22,91

Для повышения точности измерений (уменьшения невязки) времен вступлений волн необходимо использовать методы цифровой фильтрации, например, вейвлет-фильтрацию [16].

Усредненные данные таблицы 2 для всей совокупности экспериментов представлены таблице 3.

Таблица 3: Сравнение методов по усредненным результатам

Метод	Результаты тестирования	
	Количество итераций	Абсолютная невязка, м
Метод псевдообращения	34	38,2
Метод Качмажа	25	34,7

По отношению к экспериментальным данным из таблицы 3 видно, что адаптивный метод Качмажа дает более точные результаты в сопоставлении с методом псевдообращения, а также экономичнее по числу итераций.

3. Заключение

Выполнено численное моделирование алгоритмов, построенных на методах решения обратной задачи геолокации в мониторинговых системах- методах псевдообращения и адаптивного метода Качмажа. Моделирование по пространственным характеристикам приближено к реальной схеме расположения источников и приемников в районе проведения модельного эксперимента. Численное моделирование геолокации на синтетических данных обоими методами показало, что по точности и количеству итераций для достижения решения оба метода сопоставимы. В то же время метод Качмажа по трудоемкости вычислений наиболее соответствует требованиям эксперимента реального времени.

При работе с реальными сигналами и шумами точность геолокации падает в связи с влиянием на регистрацию повышенного уровня внешних транспортных шумов на сигналы, и как следствие, на возрастание погрешностей в определении времен прихода волн. Тем не менее, метод Качмажа в этих условиях показывает свои преимущества по точности и скорости сходимости.

4. Благодарности

Работа выполнена в рамках госзадания FWNM–2022–0004.

5. Литература

- [1] <https://info.sibnet.ru/article/485140>.
- [2] М.С. Хайретдинов, С. А. Авроров, Г. М. Воскобойникова, Г.Ф. Седухина. Оценивание сейсмоакустических эффектов техногенных взрывов с помощью сейсмических вибраторов. "Технологии сейсморазведки"– Новосибирск, № 2, 2012, с. 98–105
- [3] Хайретдинов М.С., Воскобойникова Г. М., Седухина Г. Ф., Ковалевский В. В., Павлов А. Ф. Изучение метеозависимой фокусировки акустических волн с помощью сейсмического вибратора и взрывов. Периодический Научно-технический журнал

- Национального ядерного центра республики Казахстан. Вестник НЯЦ РК. Курчатов, Казахстан, 2014. Вып.2. С. 129-137
- [4] Khairtdinov M.S., Voskoboynikova, G.M., Karavaev, D.A., Numerical Simulation of Acoustic Waves Propagation in an “Atmosphere–Forestland–Ground” System. Journal of Applied and Industrial Mathematics 13(1), 2019, с. 175-183
- [5] Khairtdinov M.S., O.K. Omelchenko, Sedukhina G.F., Voskoboynikova G.M. Numerical algorithms and results of experiments to determine the parameters of the borehole bottom and medium. Bull.Nov.Comp.Center Math. Model. In Geoph., 10 (2005), p.p. 35-45.
- [6] Т.Б. Яновская, Л.Н. Порохова. Обратные задачи геофизики. С.-Петербургский государственный университет. 2004. 213 с.
- [7] Д. Каханер, К. Моулер, С. Нэш. Численные методы и программное обеспечение. М.: «Мир», 2001, 575с.
- [8] Нолет Г. Сейсмическая томография // Под ред. Г.Нолета. – М.: Мир, 1990. – 416 с.
- [9] Тихонов А.П., Арсенин В.Я. Методы решения некорректных задач. // А.П. Тихонов, М.: Наука, Главная редакция физико-математической литературы, 1979, Изд. 2-е
- [10] С.И. Кабанихин, О.И. Криворотко. Сингулярное разложение в задаче об источнике. Сибирский журнал вычислительной математики. 2012. Т.15, №2. С. 205-211.
- [11] Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. // Дж. Форсайт, М.:Мир, 1980, 280 с.
- [12] Кочнев В.А. Адаптивные методы решения обратных задач геофизики : Учебное пособие. // В.А. Кочнев, Красноярский государственный университет, ВЦК СО РАН. 130 с.
- [13] Юлдашев Т. Модели распространения сейсмических волн в слоистой среде земного шара и алгоритмы определения параметров источника //отв. ред. О.М. Набиев; АН Респ. Узбекистан, НИИ «Алгоритм-Инжинеринг» . – Т.: Фан, 2011. – 232 с. – ISBN 978-9943-19-113-6
- [14] Омельченко О.К., Седухина Г.Ф., Хайретдинов М.С. Результаты моделирования алгоритмов автоматизированного определения координат забоя скважины // Материалы международной конференции «Информационные технологии и обратные задачи рационального природопользования», г. Ханты-Мансийск, 12-14 апреля 2005 г., с. 127-131.
- [15] Ливенец, А.А. Разработка и исследование системы автоматизированной локации искусственных сейсмических источников / А.А. Ливенец // Труды ИВМиМГ СО РАН. Сер. Информатика. - 2009. - Вып. 5: Материалы Пятой Азиатской международной школы-семинара "Проблемы оптимизации сложных систем". - С. 227-230.
- [16] Khairtdinov M.S., Avrorov S.A., Livenets A.A. A computational technology in seismic monitoring. Bull.Nov.Comp.Center., Math.Model. in Geoph. – 2010. – С.51–69.

Symbolic Computation in the Analysis of the Dynamics of an Orbital Gyrostat

Andrei V. Banshchikov

*Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences,
134 Lermontov St., Irkutsk, 664033, Russian Federation*

Abstract

The paper presents the possibilities of symbolic computation in the study of the dynamics of rotational motion along a circular orbit of a gyrostat in a Newtonian central field of forces. In accord with the problem of Lyapunov's stability from the equations of perturbed motion in the first approximation, in the space of the introduced parameters, the regions are found in which there is gyroscopic stabilization of unstable equilibria. The results were obtained with the help of applied software and functions of symbolic-numerical modeling of the computer algebra system *Mathematica*.

Keywords

Gyrostat, symbolic-numerical modeling, stability of equilibria, gyroscopic stabilization, software package of computer algebra.

1. Introduction

The process of solving problems on a computer in an analytical form is called symbolic computation. Symbolic computation [1] have long passed into the category of computer methods with a variety of applications (see, for example, [2]).

The classical problem of the influence of the structure of forces on the stability of the equilibria of mechanical systems [3] began to develop in the 19th century – the effect of gyroscopic stabilization was discovered. Nevertheless, the problem remains relevant: see, for example, the review in [4] among the large number of publications on this topic. With the help of symbolic computations, the present paper investigates the dynamics of an orbital gyrostat (with an arbitrary inertia ellipsoid) and the question of the possibility of gyroscopic stabilization of its unstable equilibrium positions.

The rigid body with the fixed axis of a statically and dynamically balanced flywheel rotating about that axis with a constant relative angular velocity is a gyrostat. The system moves along the circular Keplerian orbit in a central Newtonian field of forces around the gravitational center.

The stabilization of equilibria using the equations of the first approximation, that is presented here, continues and supplements the studies performed earlier for oblate and prolate axisymmetric gyrostat (see, for example, [5]).

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ bav@icc.ru (A. V. Banshchikov)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (icc-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.04

2. Construction of a Symbolic Model and Stability Conditions

By the construction of a symbolical model, one implies the obtaining of nonlinear and linearized differential equations of motion in analytical form in computer memory. The software package (SP) [6] used in this paper is designed for modeling and qualitative analysis in symbolic form of dynamic systems (in particular, the systems of interconnected absolutely rigid bodies). This applied software, the functional description and application technology of which is given in [7, 8], is a set of interactive programs executed in the interpretation mode in the computer algebra system (CAS) *Mathematica* environment.

Consider the position of equilibrium ($\dot{\alpha} = 0$, $\dot{\beta} = 0$, $\dot{\gamma} = 0$) in regard to orbital coordinate system (OCS) in general form:

$$\alpha = \alpha_0 = \text{const}, \quad \beta = \beta_0 = \text{const}, \quad \gamma = \gamma_0 = \text{const}. \quad (1)$$

Here α, β, γ are the aircraft angles that specify the orientation of the principal central axes relative to the OCS.

The equations of motion of the satellite-gyrostatt with respect to its center of mass in Euler form are quite well-known (see, for example, [9]). With the help of SP [6], the following results in an analytical form in computer memory are obtained:

- (a) kinetic energy and force function of the approximate Newtonian field of gravitation;
- (b) nonlinear differential equations in Lagrange form of the second kind describing the motion of an orbital gyrostatt;
- (c) existence conditions of equilibrium (1).

For example, Figure 1 shows the least bulky equation of motion with respect to the generalized coordinate γ (i.e. this is one of the three equations (b)).

$$\begin{aligned} & A (\ddot{\gamma} + \ddot{\alpha} \sin[\beta]) + \\ & \frac{1}{2} \left(2 \cos[\beta] (A + (-B + C) \cos[2\gamma]) \dot{\alpha} \dot{\beta} + (B - C) \cos[\beta]^2 \dot{\alpha}^2 \sin[2\gamma] + (-B + C) \dot{\beta}^2 \sin[2\gamma] \right) + \\ & \omega \left(\cos[\beta] \dot{\alpha} ((B - C) \cos[\beta] \sin[2\gamma] + \sin[\gamma] h_2 + \cos[\gamma] h_3) + \right. \\ & \quad \left. \dot{\beta} (\cos[\beta] (A + (-B + C) \cos[2\gamma]) - \cos[\gamma] h_2 + \sin[\gamma] h_3) \right) + \\ & \omega^2 \left((B - C) (3 \cos[\alpha] \cos[2\gamma] \sin[\alpha] \sin[\beta] + \cos[\gamma] (3 \cos[\alpha]^2 + \cos[\beta]^2 - 3 \sin[\alpha]^2 \sin[\beta]^2) \sin[\gamma]) + \right. \\ & \quad \left. \cos[\beta] \sin[\gamma] h_2 + \cos[\beta] \cos[\gamma] h_3 \right) = 0 \end{aligned}$$

Figure 1: The equation of motion in Lagrange form with respect to the coordinate γ .

where A, B , and C are the moments of inertia of the gyrostatt relative to its principal central axes, and h_j ($j = \overline{1, 3}$) are the projections of gyrostatic moment vector of system onto these axes. See [10] for a detailed description of the system's motion and an explicit form of the conditions (c).

Necessary conditions of stability for the equilibrium can be obtained from the equations of perturbed motion in the first approximation. The linearized equations of perturbed motion in vicinity of (1) look this way:

$$M \ddot{q} + G \dot{q} + K q = 0, \quad (2)$$

where $q = (\bar{\alpha}, \bar{\beta}, \bar{\gamma})^T$ is the column vector of deviations of generalized coordinates from the unperturbed motion (1); M is a positive definite symmetric matrix of kinetic energy; G is a skew-symmetric matrix of gyroscopic forces; and K is a symmetric matrix of potential forces.

The structure and the explicit form of the elements of the matrices M , G , K are presented in [10]. All derivatives in (2) are calculated by dimensionless time $\tau = \omega t$. Here ω is the module of orbital angular velocity.

The characteristic equation: $\det(M\lambda^2 + G\lambda + K) = v_3\lambda^6 + v_2\lambda^4 + v_1\lambda^2 + v_0 = 0$ of system (2) contains λ only in even degrees. The stability of equilibrium (1) takes place when all roots with respect to λ^2 , being simple, will be real negative numbers. The algebraic conditions providing specified properties of roots (*necessary conditions of stability*), represent the system of inequalities:

$$\begin{cases} v_3 \equiv \det M > 0, & v_2 > 0, & v_1 > 0, & v_0 \equiv \det K > 0, \\ Dis \equiv v_2^2 v_1^2 - 4v_1^3 v_3 - 4v_2^3 v_0 + 18v_3 v_2 v_1 v_0 - 27v_0^2 v_3^2 > 0. \end{cases} \quad (3)$$

Emphasize that the construction of the symbolic linearized model (2) (i.e., obtaining in the analytical form the elements of the matrices M , G , K), the calculation of the coefficients v_i ($i = \overline{0,3}$) and the discriminant Dis from (3) was also performed using the SP [6].

3. Parametric Analysis of Gyroscopic Stabilization Conditions

In [11], using the algorithms of constructing Gröbner bases, all the equilibrium positions in regard to OCS of a satellite-gyrostad are determined analytically or numerically for three special cases. For these cases, the gyrostatic moment vector is in one of the planes formed by the satellite's principal central axes of inertia. For example, in case ($h_3 = 0$, $h_1 \neq 0$, $h_2 \neq 0$), there is the equilibrium orientation:

$$\begin{cases} \alpha = \alpha_0 = \pi/2, & \gamma = \gamma_0 = 0, \\ \beta = \beta_0 = \text{const} : & h_2 \sin\beta_0 - \cos\beta_0 (h_1 + 4(A - B) \sin\beta_0) = 0. \end{cases} \quad (4)$$

Let us parametrize the problem. Without loss of generality, let $h_i > 0$, ($i = 1, 2$), and $B > A > C$ for definiteness. Let us introduce dimensionless parameters:

$$H_1 \equiv \frac{h_1}{B}; \quad H_2 \equiv \frac{h_2}{B}; \quad J_A \equiv \frac{A}{B}; \quad J_C \equiv \frac{C}{B}; \quad p_c \equiv \cos\beta_0; \quad p_s \equiv -\sin\beta_0. \quad (5)$$

The values of the parameters belong to the intervals:

$$\begin{aligned} H_i > 0, \quad (i = 1, 2); \quad & 1/2 < J_A < 1, \quad 1 - J_A < J_C < J_A; \\ & -1 < p_c < 1, \quad (p_c \neq 0, \quad p_s = \pm\sqrt{1 - p_c^2}). \end{aligned} \quad (6)$$

According to Kelvin–Chetaev's theorems [3], studying the questions on stability of equilibria begins with an analysis of the matrix of potential forces. For applied problems of spacecraft dynamics, one usually sets the distribution of masses in the system, under which the initial matrix of potential forces will be positive definite. Further, due to the influence of dissipative forces, the asymptotic stability of motion is ensured by the Lyapunov theorem. However, potentially unstable systems may also be of interest, for example, because of the possibility of nonstandard situations in orbit.

The equation from (4) in notations (5) is resolved with respect to the parameter H_1 as follows: $H_1 = p_s (4(J_A - 1) - H_2/p_c)$. Considering last relation and notations (5), the matrix of potential forces for the equilibrium (4) takes the form:

$$K = \begin{pmatrix} 3(J_C - p_s^2 - J_A p_c^2) & 0 & -3(J_C - 1)p_s \\ 0 & K_{22} & 0 \\ -3(J_C - 1)p_s & 0 & K_{33} \end{pmatrix}, \quad (7)$$

where $K_{22} = H_2/p_c + 4(1 - J_A)p_c^2$; $K_{33} = H_2 p_c + (1 - J_C)(2(p_c^2 - p_s^2) - 1)$.

It is not difficult to show that the principal diagonal first-order minor of the matrix K from (7) on the intervals (6) is negative. Hence, the matrix of potential forces is not positive definite and equilibrium (4) will be unstable.

The parameter p_s enters the coefficients of the system's characteristic equation only in even degrees. Let us eliminate it, considering $p_c^2 + p_s^2 = 1$. Let us write down these coefficients depending on four parameters J_A, J_C, p_c, H_2 in an explicit form:

$$\begin{aligned} v_3 &\equiv \det M = J_A J_C p_c^2; & v_2 &= H_2^2 (J_A - (J_A - 1)p_c^2) + \\ &+ H_2 p_c ((J_A - 1)p_c^2 (6J_A + J_C - 2) - J_A (6J_A + J_C - 7)) + \\ &+ p_c^2 \left(-(J_A - 1)p_c^2 \left(2(3J_A + 1)J_C + (1 - 3J_A)^2 - 3J_C^2 \right) + \right. \\ &\quad \left. + (J_A (3J_A - 2) - 3)J_C + 9J_A (J_A - 1)^2 + 3J_C^2 \right); \\ v_1 &= H_2^2 (J_A + 3J_C - 3 - 4(J_A - 1)p_c^2) + H_2 p_c (J_A (22 - 19J_C) - 3J_A^2 + \\ &\quad + (J_A - 1)p_c^2 (6J_A + 19J_C - 26) - 3(J_C - 7)(J_C - 1)) + \\ &\quad + p_c^2 ((J_A - 1)p_c^2 (6J_A (5 - 7J_C) + 9J_A^2 - 3J_C^2 + 34J_C - 31) + \\ &\quad + 3(J_C - 1) \left((3J_A - 2)J_C + 9(J_A - 1)^2 \right)); \\ v_0 &\equiv \det K = 3(H_2 - 4(J_A - 1)p_c^3) * \\ &* (H_2 (J_C - 1 - (J_A - 1)p_c^2) + p_c (J_C - 1) (4(J_A - 1)p_c^2 - 3J_A - J_C + 4)). \end{aligned} \quad (8)$$

It is known that if the equilibrium position is unstable at potential forces, Kelvin–Chetaev's theorem [3] of influence of gyroscopic forces tells us that gyroscopic stabilization is possible only for systems with an even degree of instability. The evenness (or oddness) of the degree of instability according to Poincare is determined by positivity (or negativity) of the determinant of the matrix of potential forces.

Let us pose the question of the possibility of gyroscopic stabilization of unstable equilibrium (4) under condition $\det K > 0$. With the help of *Mathematica* function:

$$\text{Reduce}[\{1/2 < J_A < 1, 1 - J_A < J_C < J_A, -1 < p_c < 1, p_c \neq 0, H_2 > 0, \det K > 0\}, \\ \{J_A, J_C, p_c, H_2\}, \text{Reals}]$$

designed to find the symbolic (analytical) solution of the inequalities systems, the region with an even degree of instability is obtained. Due to the solution bulkiness, its presentation is omitted here. An analysis of the solution obtained allows us to conclude the following conclusion.

Proposition 1. *The region with an even degree of instability for equilibrium (4) with the values of parameters J_A, J_C from (6) lies in the plane: $-1 < p_c < 0 \wedge 0 < H_2 < 2$.*

Next, we will use two more *Mathematica* functions *RegionPlot*, *RegionPlot3D*, designed, respectively, for a graphical 2D and 3D representations of the solution of the inequalities system.

For the detection of a property of gyroscopic stabilization, it is necessary to find in what part of region with an even degree of instability the remaining inequalities from (3) are fulfilled (except for $v_3 \equiv \det M > 0$, $v_0 \equiv \det K > 0$). It is not possible to obtain an analytical solution for the entire system of inequalities (3) because of the large number of parameters and the complexity of the expressions being analyzed. For example, the discriminant *Dis* (see (3)) of a cubic equation is an 8th degree polynomial in regard to H_2 with the coefficients depending in a complicated manner on the parameters J_A , J_C , and p_c . Their maximum degrees are 12, 8, 18, respectively. Therefore, to simplify the analysis, let us move on to symbolic-numerical modeling for fixed values of one or two parameters.

For certainty, let one of the parameters have the following value: $p_c = -\sqrt{2}/2$. Instead of solving system (3), let us first consider a simpler problem. The region of positivity of coefficients (8) was obtained using the function:

$$\text{RegionPlot3D}[1 - J_A < J_C < J_A \wedge v_0 > 0 \wedge v_1 > 0 \wedge v_2 > 0, \{J_A, 1/2, 3/4\}, \\ \{J_C, 39/100, 3/4\}, \{H_2, \sqrt{2}/2, \sqrt{2}\}]$$

and is presented in Figure 2.

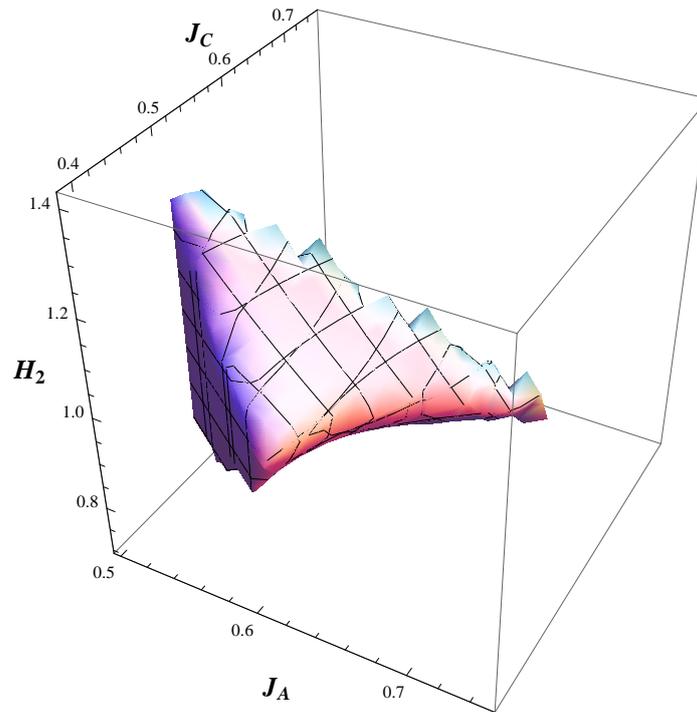


Figure 2: The region of positivity of coefficients (8) for $p_c = -\sqrt{2}/2$.

The boundary values of the parameters J_A , J_C , H_2 were in advance found by the function *Reduce* with a change in their orderliness.

Drawing a conclusion from the three-dimensional symbolic-numerical modeling, we can formulate

Proposition 2. *The gyroscopic stabilization of the unstable equilibrium (4) is possible only for values of the parameters (5) from the intervals:*

$$1/2 < J_A < 3/4 \wedge 1 - J_A < J_C < J_A \wedge -1 < p_c < -1/2 \wedge 1/2 < H_2 < 2.$$

Experience shows that three-dimensional images are often complex and not informative enough to draw conclusions on. Therefore, further parametric analysis of the system of inequalities (3) was carried out with respect to two selected parameters.

In addition to the value of the parameter p_c , let the second parameter have the value: $J_C = 1/2$. Let us construct the regions with an even degree of instability and of gyroscopic stabilization in the parameter plane J_A, H_2 using function

$$\text{RegionPlot}[1/2 < J_A < 1 \wedge 0 < H_2 < 2 \wedge v_0 > 0 \wedge v_1 > 0 \wedge v_2 > 0 \wedge Dis > 0, \\ \{J_A, 1/2, 1\}, \{H_2, 0, 2\}].$$

The result obtained is shown with colour regions in Figure 3.

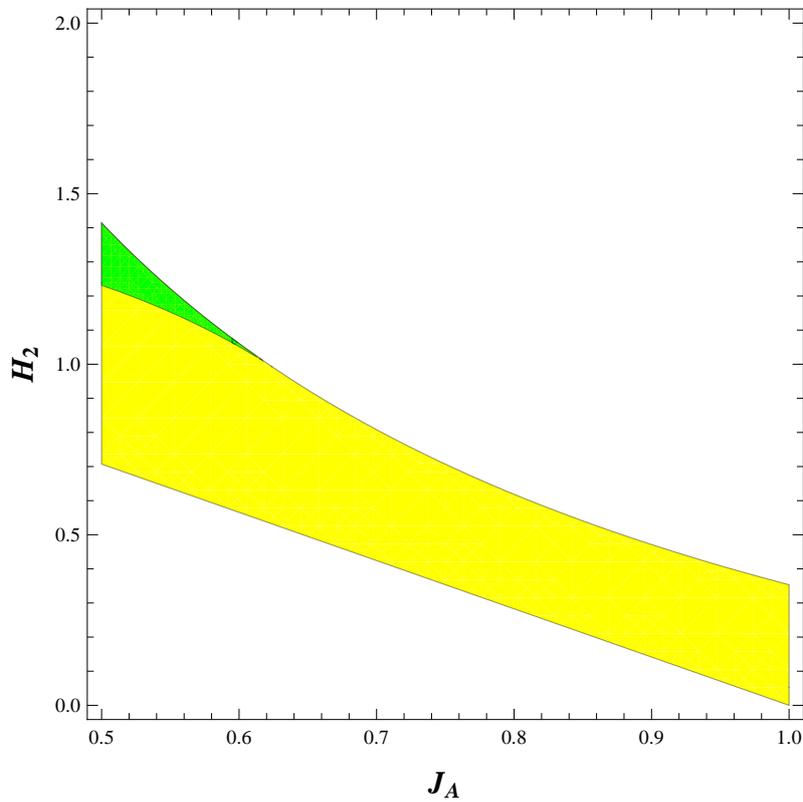


Figure 3: The found areas for $p_c = -\sqrt{2}/2$, $J_C = 1/2$.

The yellow area in the figure is the region with an even degree of instability. Its green part determines the parameter values at which gyroscopic stabilization is possible. Outside the

selected regions, the system has an odd degree of instability (i.e. $\det K < 0$) and equilibrium (4) is unstable.

4. Conclusion

By analogy with the parametric analysis presented above, the possibility of gyroscopic stabilization for other equilibrium positions was also examined. It is important to note that this question is not always resolved positively (see, for example, [10]).

The presented results of the dynamics of an orbital gyrostat using symbolic-numerical calculations indicate that the proposed approach expands our capabilities in the study of multi-parameter problems.

References

- [1] B. Buchberger, G. E. Collins, R. Loos, R. Albrecht, *Computer Algebra: Symbolic and Algebraic Computation*, Springer-Verlag, New York, 1983.
- [2] J. Grabmeier, E. Kaltofen, V. Weispfenning, *Handbook in Computer Algebra. Foundations, Applications, Systems*, Springer-Verlag, Berlin, 2003.
- [3] N. G. Chetaev, *The Stability of Motion*, Pergamon Press, New York, 1961.
- [4] S. A. Agafonov, The stability and stabilization of the motion of non-conservative mechanical systems, *J. Appl. Math. Mech.* 74(4) (2010) 401–405.
- [5] A. V. Banshchikov, S. V. Chaikin, Analysis of the stability of relative equilibria of a prolate axisymmetric gyrostat by symbolic-numerical modeling, *Cosm. Res.* 53(5) (2015) 378–384.
- [6] A. V. Banshchikov, V. D. Irtegov, T. N. Titorenko, Software package for modeling in symbolic form of mechanical systems and electrical circuits, 2016. Certificate of State Registration of Computer Software No. 2016618253, Federal service for intellectual property, Issued July 25th., 2016. (in Russian).
- [7] A. V. Banshchikov, L. A. Burlakova, V. D. Irtegov, T. N. Titorenko, Symbolic computation in modelling and qualitative analysis of dynamic systems, *Comput. Technol.* 19(6) (2014) 3–18. (in Russian).
- [8] A. V. Banshchikov, A. A. Vetrov, Application of software tools for symbolic description and modeling of mechanical systems, in: I. V. Bychkov et al. (Eds.), *CEUR Workshop Proceedings of the 2nd. International Workshop on Information, Computation, and Control Systems for Distributed Environments*, 2020, pp. 33–42. URL: <http://ceur-ws.org/Vol-2638/paper3.pdf>.
- [9] V. A. Sarychev, S. A. Mirer, A. A. Degtyarev, Dynamics of a gyrostat satellite with the vector of gyrostatic moment in the principal plane of inertia, *Cosm. Res.* 46(1) (2008) 60–73.
- [10] A. V. Banshchikov, Symbolic-numerical analysis of the necessary stability conditions for the relative equilibria of an orbital gyrostat, *J. Appl. Indust. Math.* 14(2) (2020) 213–221. doi:10.1134/S1990478920020015.
- [11] S. A. Gutnik, V. A. Sarychev, Application of computer algebra methods for investigation of stationary motions of a gyrostat satellite, *Program. Comput. Softw.* 43(2) (2017) 90–97. URL: <https://doi.org/10.1134/S0361768817020050>.

Hierarchical Control Development Environment: Requirements and Cases

Evgenii A. Cherkashin^{1,2}, Nadezhda V. Nagul¹, Yingjie Wang³ and Igor V. Bychkov^{1,3}

¹*Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences, 134 Lermontov St, Irkutsk, 664033, Russian Federation*

²*Irkutsk State University, Karl Marx St, 1, Irkutsk, Irkutsk Oblast, 664003, Russian Federation*

³*Yantai University, 32 Qingquan Avenue, Yantai, Shandong 264005, China*

Abstract

The paper presents an approach to solve the problem of reducing complexity of representing hierarchical control systems of autonomous underwater vehicles (AUV), their models and interaction with external objects, as well as hardware in the form of information objects. For this purpose, a specialized environment for research is being developed, including a conceptual model of the subject area in the form of a formalized ontology, a domain specific language (DSL), a translator from this language, a system for transforming the obtained structures into software objects and data objects of other programming systems, thus providing the upper level of the software management.

Keywords

hierarchical control, intelligent technologies, intelligence and cognitive services, model driven architecture, autonomous underwater vehicles, convolutional neural network control

1. Introduction

The use of a group of Autonomous Underwater Vehicles (AUVs) has many advantages over a single robot, one of which is the ability to perform complex tasks such as ocean patrol and exploration, resource exploration and environmental monitoring. Thanks to the cooperation and information sharing between individual team members, an AUV group is more reliable and can adapt to different conditions and requirements. AUV groups can achieve more efficient resource utilization and mission accomplishment through distributed deployment and cooperative management, thereby improving the quality of research results. However, as the size of a group of AUVs increases due to the uncertainty in the parameters of each group member and the perturbations associated with its motion, more stringent requirements are imposed on the control strategy of AUVs in terms of accuracy, reliability, stability, etc. To solve the problems of controlling AUV formations under unknown and uncertain conditions, methods that are general enough to effectively deal with various group control problems have not yet been developed,

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ eugeneai@icc.ru (E. A. Cherkashin); sapling@icc.ru (N. V. Nagul); wangyingjie@ytu.edu.cn (Y. Wang); bychkov@icc.ru (I. V. Bychkov)

🆔 0000-0003-2428-2471 (E. A. Cherkashin); 0000-0003-1439-3274 (N. V. Nagul); 0000-0001-5236-3565 (Y. Wang); 0000-0002-1765-0769 (I. V. Bychkov)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (icc-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.05

and there is a lack of complete research software to implement these methods. The methods of automatic synthesis of hierarchical control of complex formations of AUVs are needed, which take into account individual peculiarities of AUV functioning, provide joint solution of tasks by a group of robots, limited acoustic communication channel, small choice of environmental sensors, partial or complete absence of a map of the work area, a priori uncertainty and variability of functioning conditions, inaccuracy of information received from measuring devices, possibly variable composition of the AUV grouping, significant limitations of the grouping of AUVs, and other factors. In this connection, the development of hierarchical control means is an urgent task of controlling heterogeneous (in terms of variants of control realization in a particular object) AUVs, especially in a dynamic environment with uncertainties. Deeper automation of AUV group control can be achieved through the use of artificial intelligence (AI) technologies. The paper proposes a way of developing hierarchical control systems, where the final control is built on the basis of discrete event model analysis, including lower level controls, and represents synthesized adaptive control strategies describing group behavior in the form of scenarios.

2. Problem statement

Currently, there is a steady tendency of increasing scientific interest in decentralized network-communication control of multi-agent systems. The most popular subject area of articles published on the subject can be named the research of various approaches to the construction of action planning systems for groups of autonomous robots performing survey and search operations, patrol missions and inspection tasks of various objects and structures. One of the most actively researched problems in the field of multi-agent systems is the control of formations of moving agents. More and more attention is paid to the tasks of controlling formations of aircraft [1], autonomous ground-based [2] and underwater robots [3]. Moreover, if in early studies more attention was paid to formations of aircraft and land-based robots, now the emphasis is shifting towards autonomous underwater robots, which is explained by the need to explore the largely unexplored depths of the World Ocean and its resources [4]. There is also a tendency to complicate the formulation of formation control problems: formulations with nonholonomic models of agents, nonlinear links between them, and variable topology of the information network are more and more often considered. Regardless of the approaches used, the following main directions of the current research in the field of AUVs formation control can be emphasized: development of robust and adaptive control systems for a group of AUVs taking into account the specifics of the underwater environment, e.g. the effect of lag and possible information loss during data transfer between AUVs, as well as the use of groups of AUVs in performing survey, search and protection missions.

Considering the dynamism and uncertainty of real-life environments and operating conditions, the research and development of autonomous control systems for mobile robots includes algorithmic control, PID control, neural network-based control, etc. Lower control levels respond to events generated by the upper levels, and switch in accordance with the main control goals. It is necessary to develop tools for recognizing the states of the system at the upper levels of control and implement control synthesis for the lower levels using, for example, logical inference and large language models. Lower-level control strategies will be determined by upper-

level decisions in the form of specific behavioral scenarios, including path planning, generation of input data for pre-trained neural networks (graph convolution), control based on autonomous evolutionary games, and specific control signals for actuators. Therefore, it is needed to develop a multi-level hierarchical control system for a group of AUVs based on distributed control technology of event-triggered models, symbolic information processing and logical inference. To increase the stability of the system in an uncertain and complex environment and the accuracy of issued commands a switching control modes are to be implemented in accordance with control goals.

The path planning problem of a group of AUVs based on the hierarchical control requires developing appropriate distributed control strategies and algorithms by decomposing the path planning problem into several levels in order to realize flexible path planning and decision making for groups of AUVs at different levels. The development goals are to optimize global trajectory planning, reduce time to complete tasks, accuracy of local trajectory planning, implement dynamic obstacle avoidance with the collection of information about the environment, which will lead to increased autonomy and adaptability of AUVs in difficult conditions, increase the efficiency and autonomy of control and the quality of distribution tasks in the AUV group.

3. Approaches suggested

3.1. Discrete-event systems

To solve the problems arising in group control of AUVs, we propose to use an approach based on discrete-event systems (DES), in which the change of the system state occurs due to the occurrence of some events of various nature and character. In a hierarchical control structure, the system and its associated lower level controllers can be represented as DES for the upper control level.

Hundreds of journal publications and dozens of monographs are devoted to analytical studies of DESs. A detailed overview of the current state and trends in DES theory can be found in recent works [5], [6], [7]. Existing methods of DES study reflect the diversity of their models: finite automata theory, Petri nets, minimax algebras, Markov chains, and others. It was proposed a theoretical basis of a generalized approach to the study of DESs, which is potentially equally effective for any of the above models, based on the generation of a language (an infinite sequence of states and events) generated on the basis of the DES model. In doing so, the properties of formal regular languages are analyzed. In the specific cases, AUVs control may be represented as DES of low level and be a part of synthesized hierarchical control.

To control DESs containing events that can be prohibited from occurrence, the supervisory control theory (SCT) is used. The SCT was developed as a tool for constraining the behavior of the DES in accordance with a set of constraints defined by some formal specification. A detailed description of the SCT, outlining the results obtained in recent years, can be found in [8],[9]. The traditional area of application of DESs and SCT is the sphere of production processes, however, intensive research is currently being carried out on the features of the application of SCT in the field of robotics and robots group control. It seems reasonable to generalize the approach of analyzing DES properties in the direction of applying universal methods of mathematical systems' theory. For example, the reduction method proposed by S.N. Vassiliev

and the method of logical-algebraic equations developed in ISDCT SB RAS may be applied. As a system for automating the above methods, the calculus of positively constructed formulas (PCFs) and the method for automatic theorem proving (ATP) developed on its basis may be used. The PCF calculus advantageously differs from the capabilities of other logical means of formalizing the subject area and searching for logical conclusions: expressiveness combined with compactness of knowledge representation, "natural" parallelism of its processing, large-block and less combinatorial complexity of inferences, high compatibility with heuristics. Previously, the basic problems of SCT including controllability checking, supremal controllable sublanguage of a given specification language construction and a monolithic supervisor realization were successfully solved using PCF-based approach [10]. Also, partially observed DES were studied, in which occurrence of some events are unavailable for observation [11].

3.2. Domain specific languages

Development of domain specific languages (DSL) is a popular direction in computer science of solving problems of abstract representation of a modeled object in terms close in level to the problem and methods of its solution. The main problem to be solved is the adaptation of the problem to the tools for its solution. In our case, we need a tool to describe different levels of DES, methods of control synthesis in terms of control theory, combinations of search strategies, physical models of AUVs, connection with external components (configurations), as well as assignments of specific DES of lower levels of control in the form of automata, transition rules, etc. Adaptation to tools is realized by methods of generative programming (Czernecki, Gorbunov-Posadov) with the help of various tools: compiler compilers, template programming, Model-Driven Architecture, MDA.

The potential of using MDA (Model Driven Architecture, a transformational approach to the design of software systems) in supporting complex scientific research is currently significantly underestimated. For example, in [12], an analysis of applications in the design of machine learning software systems shows that: the ideas of an abstract representation of data processing pipelines are not implemented to the extent that would significantly increase the productivity of research in machine learning, even less development of modeling and transformation of requirements for the pipeline. integration, automated deployment, monitoring. Existing solutions require greater maturity and better tool support; more attention must be paid to the scalability of solutions generated based on transformation. This problem arises due to the fact that the point of view on MDA is limited to the transformation of 1-to-1 models (one structure of the source model into one structure of the target) according to a certain set of rules (ATL, VRML technologies). We view the transformation process as a 1-to-n, n-to-m multi-level hierarchical process that models the human design process. Transformation is naturally represented as objects in the object-oriented logical language Logtalk [13].

To specify a language, the following models must be defined: syntax, semantics, and pragmatics. Syntax is modeled using grammars and translator implementation tools (flex, yacc; ANTLR4, and the like). Semantics is built on the basis of ontological modeling of the subject domain, using ontologies (Semantic Web) to represent the resulting conceptual models. The pragmatics of the language is realized in the algorithmic parts of the translators' description, and also often by means of a subsequent transformation procedure of some abstract intermediate representation,

e.g. by applying MDA. MDA for a particular task is implemented as a procedure for generating derived information objects as a result of analyzing the original abstract information objects. The developed methods of MDA-transformation realizations also use logical inference. The main difference and advantage of the developed methods and technologies is the rigor of the results, the guaranteed nature of the obtained judgments and estimates, and, on the other hand, constructiveness, the possibility to bring the research "to the number".

It is important that the tasks under consideration are formulated directly in terms of engineering requirements to the system (direct specifications of dynamic quality indicators), and the results are also obtained in this form. Once the models under study have been reduced to a standard form, subsequent constructions and calculations will be performed to a large extent automatically, thus shielding from the researcher the special, rather mathematically complex methods used. The following techniques relate to the main principle of control synthesis for individual AUVs and their groups for specific types of AUV.

3.3. Convolutional neural network control

To solve the problem of planning the distribution of tasks between AUVs in complex multitasking environments, it may be proposed to create an integrated decision support system based on a graph convolutional neural network (GCN) enforced by classical methods of automatic logical reasoning e.g., in the PCF-calculus. It will be aimed at analyzing and predicting the optimal fit between AUVs and existing tasks, supporting for real-time interactive interaction with AUV control and planning systems, intuitive task monitoring, parameter tuning and visualization of behavioral patterns, providing effective and consistent decision support. To improve the tools for analyzing DES, implemented using logical means of inferring PCFs, the possibility of using convolutional neural networks in the heuristic apparatus for refuting PCFs should to be also explored.

The convolutional neural network control is implemented using standard steps: a) data preparation and preprocessing: collection of historical data on the results of AUV problem-solving execution, including task type, completion status, environmental parameters, etc.; b) construction of a GCN model for AUV and task conformance management to predict the best conformance to the management task plan at the AUV group level; c) the resulting plan is formally verified for compliance with task requirements and available resources using our knowledge-based system (KBS) based on PCFs; d) during plan execution, the KBS verifies compliance with constraints, task population structure. e) analyzing the obtained post-factum information to train further the neural network [14].

4. Path planning

The AUV path planning is realized at two levels: a) upper (global), where the AUV movement trajectories are designed, and b) lower (local), where new data obtained during the movement along a given path are taken into account. The upper level is constructed taking into account the task distribution, the space of possible states, and the environment constraints. It is supposed that the path is built with the help of logical prediction of the AUV motion in virtual time. To represent the model in the toolboxes, the corresponding DSL section is implemented. Global

planning is realized by adapting the modern RRT method, using random sampling and fast exploration tree methods, obstacle avoidance algorithm, e.g. artificial potential field or dynamic window method. The resulting paths are locally optimized to match the local data. The planning algorithms are iterated into the overall hierarchical control mechanism in the form of intermediate levels. The problem of role allocation in a heterogeneous group of AUVs is solved using the method of modeling the behavioral revenue function in multi-agent systems organized in multilevel structures, which should have the property of consistency of the problem specification, and have prospectivity estimates close to the optimal ones.

The planning is implemented by means of the following steps: a) strategy derivation in a multilevel joint evolutionary game: conduct a multilateral multi-strategy evolutionary game in order to obtain a Nash equilibrium strategy; b) strategy evaluation: conduct a formal evaluation of the game results, compare individual AUV behavior plans and implement the adjustment of the game rules according to the results, i.e. adjust the game model to obtain results close to optimal. On the basis of the obtained evolutionary game models, the planning of AUV allocation to tasks is realized depending on the specific specifications of the task and AUVs. The logic of simulating the behavior between groups of AUVs allows using the methods of replicator dynamics, strategic balance points in the autonomous evolutionary game model. The fundamental idea is the adaptation of abstract control theory methods to the existing low level control techniques developed earlier; developing a new DSL capable to express various aspects of the domain and methods, transformation procedures for interpreting abstract language constructions in terms of specific case; development of parameter adjustment techniques for the existing techniques of control being a part of a hierarchical control.

5. Conclusion

In this paper we briefly described the idea of developing a tool for adaptation of control methods realizing different stages of solving control problems, into one common flexible mechanism of hierarchical control of groups of heterogeneous AUVs accordingly to a given task specification. The results obtained will significantly increase the efficiency of AUV use due to the implementation of learning and adaptation, synthesis and verification mechanisms when performing complex search and survey missions and joint problem-solving. The methods and means of synthesizing multilevel control developed will be based on the application of logical inference, where the process of generating information structures representing a particular class of control is described in the form of a first-order logical theory. The means of searching for logical conclusions in the calculus of positive constructed formulas (PCF) developed in Matrosov Institute for System Dynamics and Control Theory of Siberian branch of Russian Academy of Sciences, oriented to special classes of control models, having high compatibility with heuristics, allow eliminating the known problems of the logical approach to knowledge representation and processing, including in application to the construction of the upper level of control systems. In addition, the available means of logical inference are combined with models of problem-solving, resource allocation and control of constraints and parameters in the execution of plans, decision-making on control reconfiguration, and verification of results with respect to the task specification.

References

- [1] K. Telli, O. Kraa, Y. Himeur, A. Ouamane, M. Boumehraz, S. Atalla, W. Mansoor, A comprehensive review of recent research trends on unmanned aerial vehicles (uavs), *Systems* 11 (2023). URL: <https://www.mdpi.com/2079-8954/11/8/400>. doi:10.3390/systems11080400.
- [2] A. Soni, H. Hu, Formation control for a fleet of autonomous ground vehicles: A survey, *Robotics* 7 (2018). URL: <https://www.mdpi.com/2218-6581/7/4/67>. doi:10.3390/robotics7040067.
- [3] J. Hwang, N. Bose, S. Fan, Auv adaptive sampling methods: A review, *Applied Sciences* 9 (2019). URL: <https://www.mdpi.com/2076-3417/9/15/3145>. doi:10.3390/app9153145.
- [4] B. Hadi, A. Khosravi, P. Sarhadi, A review of the path planning and formation control for multiple autonomous underwater vehicles, *Journal of Intelligent Robotic Systems* 101 (2021). doi:10.1007/s10846-021-01330-4.
- [5] C. G. Cassandras, S. Lafortune, *Introduction to Discrete Event Systems*, Springer Cham, 2021. doi:10.1007/978-3-030-72274-6.
- [6] S. Lafortune, Discrete event systems: Modeling, observation, and control, *Annual Review of Control, Robotics, and Autonomous Systems* 2 (2019). doi:10.1146/annurev-control-053018-023659.
- [7] C. Seatzu, M. Silva, J. H. van Schuppen (Eds.), *Control of discrete-event systems*, Springer London, 2013. doi:10.1007/978-1-4471-4276-8.
- [8] W. M. Wonham, K. Cai, *Supervisory Control of Discrete-Event Systems*, Springer International Publishing, 2019.
- [9] W. Wonham, K. Cai, K. Rudie, Supervisory control of discrete-event systems: A brief history, *Annual Reviews in Control* 45 (2018) 250 – 256. URL: <http://www.sciencedirect.com/science/article/pii/S1367578817301876>. doi:<https://doi.org/10.1016/j.arcontrol.2018.03.002>.
- [10] A. Davydov, A. Larionov, N. Nagul, On the application of the calculus of positively constructed formulas for the study of controlled discrete-event systems (in russ.), *Modeling and Analysis of Information Systems* 31 (2024) 54–77. doi:10.18255/1818-1015-2024-1-54-77.
- [11] A. Davydov, A. Larionov, N. Nagul, Analysis and control of partially observed discrete-event systems via positively constructed formulas, *Computation* 12 (2024). URL: <https://www.mdpi.com/2079-3197/12/5/95>. doi:10.3390/computation12050095.
- [12] H. Naveed, C. Arora, H. Khalajzadeh, J. Grundy, O. Haggag, Model driven engineering for machine learning components: A systematic literature review, *Information and Software Technology* 169 (2024) 107423. URL: <https://www.sciencedirect.com/science/article/pii/S0950584924000284>. doi:<https://doi.org/10.1016/j.infsof.2024.107423>.
- [13] E. A. Cherkashin, A. O. Shigarov, V. V. Paramonov, Representation of mda transformation with logical objects, 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (2019) 0913–0918. URL: <https://api.semanticscholar.org/CorpusID:210697961>.
- [14] Z. Zhan, Y. Wang, P. Duan, A. M. V. V. Sai, Z. Liu, C. Xiang, X. Tong, W. Wang, Z. Cai, Enhancing worker recruitment in collaborative mobile crowdsourcing: A graph neural

network trust evaluation approach, IEEE Transactions on Mobile Computing (2024) 1–18.
doi:10.1109/TMC.2024.3373469.

Usage of Threat Modelling Tools for Implementation of ICT Security Standards in IoT Environments

Dalibor Dobrilovic¹, and Rade Dragović²

¹ University of Novi Sad, Technical Faculty “Mihajlo Pupin” Đure Đakovića b.b., Zrenjanin, 23000, Serbia

² Institute for standards and technology, Resavska 28, Belgrade, 11000, Serbia

Abstract

The development of microcontrollers, communication, and sensor technologies resulted with the growing implementation of the Internet of Things (IoT) in various environments. The nature of IoT assumes the complexity of the system. The complexity of the system increases the number of system elements and therefore increases the system vulnerabilities. As a result, it is important to find an efficient and easy-to-use methodology for implementing security standards in newly designed IoT systems. This paper introduces the usage of threat modeling tools for modeling IoT systems and identifying the threats for the modeled systems. The identified threats are further embedded in the framework designed to define and enable the implementation of IoT security standards. The framework is demonstrated in the example of a simple IoT system designed for urban pollution, and noise monitoring.

Keywords

Threat modeling, threat modeling tools, Risk assessment, IoT security implementation framework, ISO security standards implementation, Internet of Things (IoT)

1. Introduction

The growing implementation of the Internet of Things (IoT) in various environments increases the general ICT security risks. The IoT environments being heavily involved in the life of users of its benefits greatly increase the security risks for the individuals using those systems or their services. The complexity of the system and the number of its components and devices give numerous possibilities of system misuse, data compression, and other forms of threats. Therefore, it is highly important to define an easy-to-use and efficient framework for identifying threats and defining the implementation of security standards within the process of designing the system.

This paper introduces the framework based on the usage of threat modeling tools for modeling IoT systems and identifying the threats for the modeled systems. The identified threats are further analyzed to define the implementation of IoT security standards. The methodology is presented on the example of a simple IoT system designed for urban pollution, and noise monitoring.

This paper is structured as follows. After the Introduction and Related Works sections, the modeled IoT system architecture is described, followed by threat modeling and threat analyses. The approach in risk assessment is then presented using the results and reports of threat modeling, thus explaining the proposed framework. At the end, the conclusions are given.

2. Related Work

According to the Open Worldwide Application Security Project (OWASP) [1] the threat modelling is the process taken to identify, communicate, and understand threats and mitigations within the context of protecting something of value. A threat model is a structured representation of all the information

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: dalibor.dobrilovic@tfzr.rs (A. 1); rade.dragovic@instate.biz (A. 2);

ORCID: 0000-0002-3083-5725 (A. 1); 0000-0002-0215-4076 (A. 2);



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.06

that affects the security of an application. There is a variety of threat modeling methods such as: STRIDE, Process for Attack Simulation and Threat Analysis (PASTA), Trike, Visual, Agile, and Simple Threat (VAST), Attack Trees, Common Vulnerability Scoring System (CVSS), Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE), Quantitative Threat Modeling Method (QTTM), DREAD, MITRE, LINNDUN etc. [2] The STRIDE, a method named on six security threats categories (Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, and Elevation of privileges) is the most popular and widely supported by the modeling tools.

In the following text will be presented several works in the area of IoT threat modeling. In [3] the authors utilize the STRIDE threat modeling methodology and the Microsoft Threat Modeling Tool to identify threats present in smart city infrastructure. Authors map offenses, possible evidence sources, and types of threats identified to help investigators understand what crimes could have been committed.

In [4] authors propose an approach aimed at supporting the security analysis of an IoT system with an almost completely automated process for threat modeling and risk assessment. The evaluation of its effectiveness is given for the application of a home automation system. The proposed methodology relies upon a modeling approach of both the architectural components of an IoT system and its security properties.

The contribution presented in [5] identifies and describes current security threats in IoT based on a generic IoT architecture and the main communication protocols that are used in the application, transport, network, and physical layer.

In the paper [6], a general-purpose methodology for assessing the risk is proposed for end-to-end systems. The approach covers static and dynamic features/components of an IoT system. The presented solution is evaluated within the real prototype implementation.

A similar approach for modeling IoT system threats is applied to the threat modeling on the container ecosystem. STRIDE is used as a threat modeling framework. [7]

3. IoT System Architecture

In this section, the general architecture of IoT systems will be described. IoT systems commonly can be presented as multi-layer systems, and the five-layer architecture of IoT systems is one of the most accepted [8]. The IoT system consists of the following layers.

The **Perception Layer** has the same functions as the Physical Layer of the OSI reference model. Sensor nodes or end devices are deployed on this layer. Sensors collect data, i.e. physical parameters of the environment. **The Network Layer** (or the Transport Layer) is designed to establish a connection to the core of the system and its components such as network devices and to transfer data collected by sensors. It is responsible for transferring sensor data from the perception layer to the Processing Layer and vice versa. For this purpose, mainly short and medium-range wireless technologies are used.

The **Processing Layer** or **Middleware Layer** is responsible for analyzing and processing collected data, storing it, and creating reports. This layer includes various technologies such as relational and non-relational databases, cloud computing, big data, etc. The application layer is responsible for delivering data to users and their visualization through user applications and devices. Finally, **The Business Layer** is on top of the system and provides the information needed for system management, business and profit model development, and data protection.

An urban IoT system designed as a sensor network for monitoring traffic noise and meteorological parameters in an urban environment [9] is presented in this section and used for further modeling and framework implementation. Figure 1 a) shows the devices of the first two layers of the IoT system. On the perception layer (I) there are sensor devices (1) for air pollution, weather conditions, and noise monitoring, Wi-Fi access point (AP) devices (2) for connecting sensor nodes with short-range Wi-Fi technology, and Wi-Fi Fi/LoRa gateway devices (3) for connecting (collecting) Wi-Fi APs in a wider area and connecting to the cloud part of the system using longer-range connection technology. Communication components use a combination of TCP/MQTT and LoRa protocols.

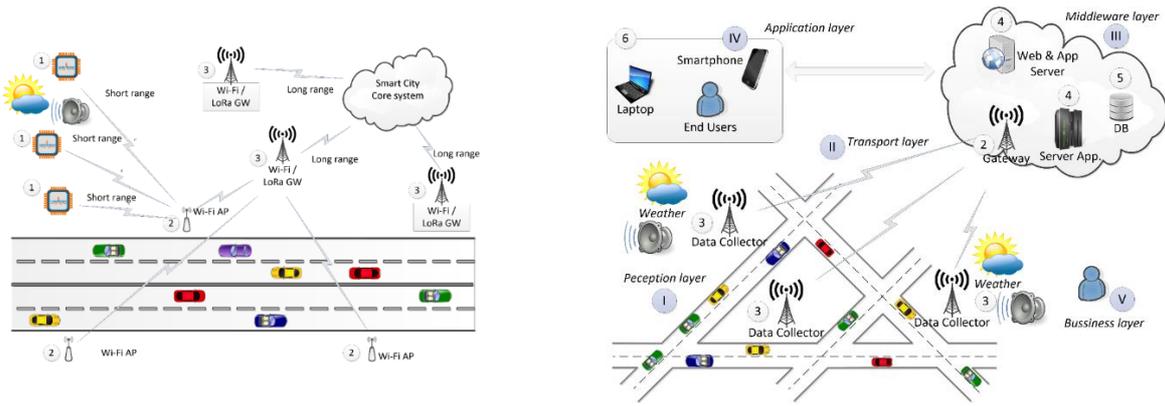


Figure 1: a) the IoT urban monitor system peripheral elements; b) layers of IoT urban monitoring IoT system

The middleware layer (III) contains a server (4) that can be installed on several servers (e.g. Data Analyzes and Application Server) or in a simplified version on one computer and a database (5). The application layer contains APIs for user applications (6) for various devices (mobile phones, laptops, desktops, etc.). The whole system is shown in Figure 1 b).

4. IoT threat modeling

Microsoft Threat Modeling Tool (MTMT) is used for described IoT threat modeling. The modeled system is shown in Fig. 2 a) in a simplified version. The elements of the system are labeled in the same way as in Fig. 1 b) with Roman (I, II, III, IV) and Arabic numerals (1, 2, 3, 4, 5, 6).

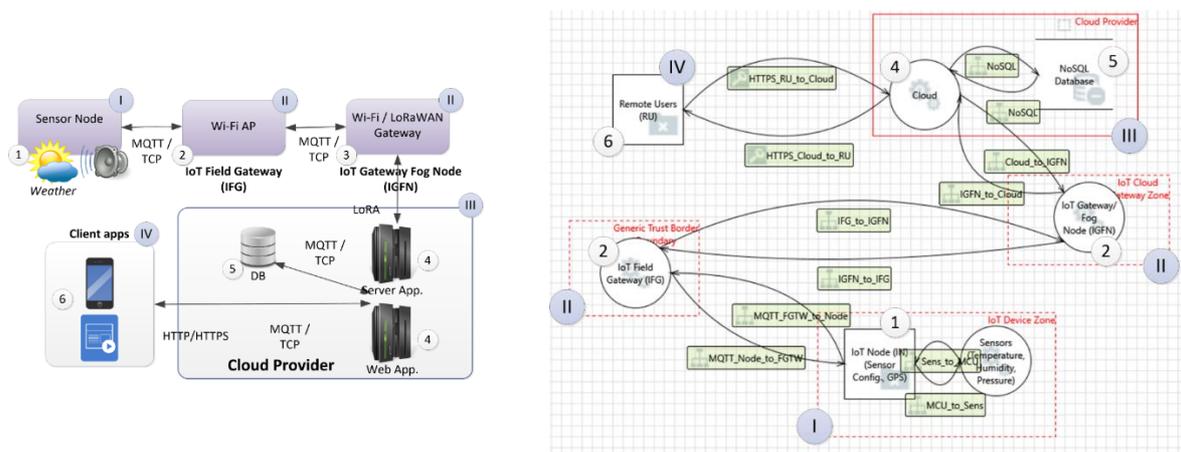


Figure 2: a) Modeled IoT system; b) Model of the system in threat modeling tool.

To model the sensor nodes (1), the *Generic External Interactor* is used in MTMT from the Stencils palette. Parameters can be set in the Element Properties palette as is shown in Fig. 3. If other devices in the External Interactor list correspond to the system being modeled, e.g. mobile phone (Smart Phone) or Browser, these devices can also be selected from the same stencil.

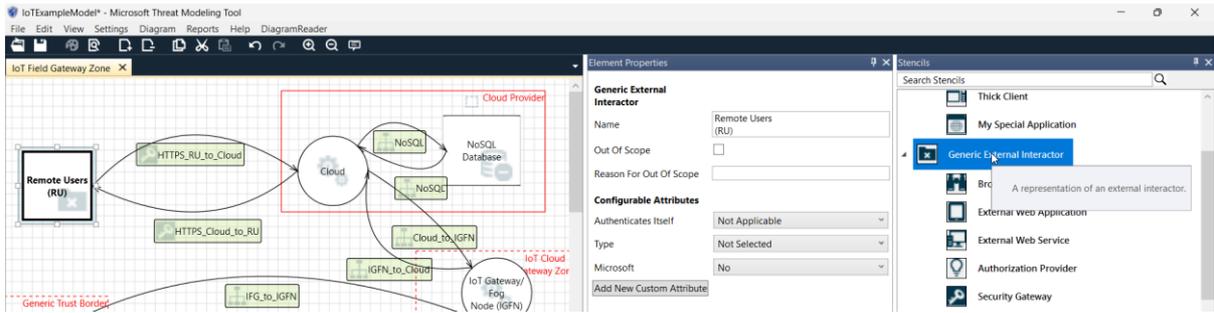


Figure 3. Use of the Interactor object

To model the Wi-Fi AP (2) and the Wi-Fi/LoRa Gateway device (3) the *Generic Process* is used (Fig. 4.). The *Generic Process* is also used to model the server (4), while the *Generic Data Store* is used to model the Database (5).

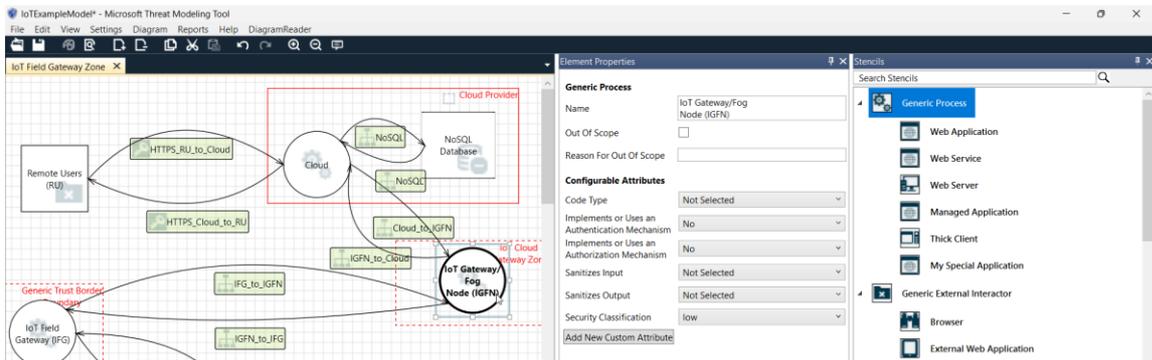


Figure 4. Use of the Process object

To model the data flow, a *Generic Data Flow* is used. An example of the HTTP protocol is shown in Fig. 5. The same element can be used for MQTT or another protocol and to model the data flow between the sensor and the MCU (microcontroller). Since the sensors are directly connected to the MCU there is no issue of insecure transmission, so the data flow can be checked as *Out of Scope* in the Element Properties palette.

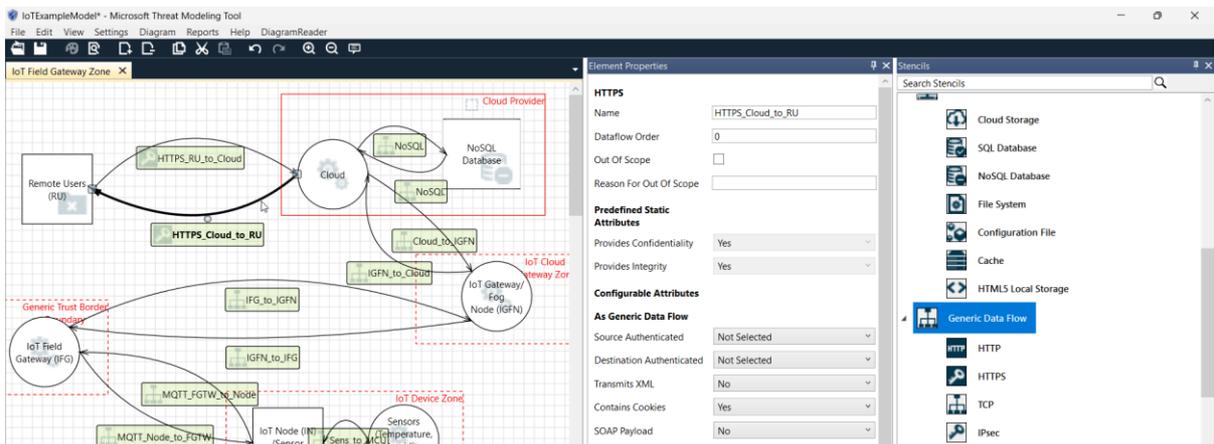


Figure 5. Using the Data Flow object for the HTTP protocol

Finally, modeling confidence limits, i.e. of separate system elements in which data is protected more than the external space, is done with the *Generic Trust Boundary* object in the Stencils palette, and it is represented in figures with dashed red lines.

4.1. Threat analysis

Threat analysis is performed by clicking on the menu option *Reports > Create Full Report*. To create a report, it is necessary to activate the *Generate Report* option, after which a report with the extension .htm will be created. The part of the report is the threat list, which can be exported to CSV format using the *Switch To Analysis View* option (Table 1).

Table 1

Partial view on exported threat list

Id	Title	Category	Interaction	Priority	Description
47	Elevation Using Impersonation	Elevation Of Privilege	IGFN_to_IFG	High	IoT Field Gateway (IFG) may be able to impersonate the context of IoT Gateway/Fog Node (IGFN) in order to gain additional privilege.
46	Elevation Using Impersonation	Elevation Of Privilege	IFG_to_IGFN	High	IoT Gateway/Fog Node (IGFN) may be able to impersonate the context of IoT Field Gateway (IFG) in order to gain additional privilege.
45	Elevation Using Impersonation	Elevation Of Privilege	Cloud_to_IGFN	High	IoT Gateway/Fog Node (IGFN) may be able to impersonate the context of Cloud in order to gain additional privilege.
44	Elevation Using Impersonation	Elevation Of Privilege	IGFN_to_Cloud	High	Cloud may be able to impersonate the context of IoT Gateway/Fog Node (IGFN) in order to gain additional privilege.
16	Elevation Using Impersonation	Elevation Of Privilege	MQTT_FGTW_to_Node	High	IoT Field Gateway (IFG) may be able to impersonate the context of IoT Node (IN) (Sensor Config., GPS) in order to gain additional privilege.

Threat analysis for the modeled generic IoT system gives the following identified threats classified into six categories (Table 2).

Table 2

The number of threats by category

No.	Threat	Count
1	Spoofing	7
2	Tampering	6
3	Repudiation	8
4	Information Disclosure	6
5	Denial Of Service	6
6	Elevation Of Privilege	12
	Total	45

5. Risk assessment

In this section, we will describe the basics of risk assessment. The basic steps are to identify Score, Context, and Criteria as a basis for establishing a framework for the information security management system. A system of early identification of threats and vulnerabilities must be established, and it must react predictively and not preventively. All the time, it is necessary to follow all identified interested parties, both those who have good intentions for the organization's system and those who have different intentions and initiatives.

Risk assessment is the next phase which must consist of risk identification, risk analysis, and risk evaluation. Risk identification is a detailed and long-term job that must identify each and every individual risk, regardless of whether it represents a higher or lower level of risk at the time of identification. This step uses the input from threat modeling explained in the previous section.

The professional framework for risk analysis is represented by the following standards:

- *ISO/IEC 27005* - Information security, cybersecurity, and privacy protection - Guidance on managing information security risks and
- *ISO 31000* - Risk management - Guidelines.

Risk analysis must consider all identified risks from the aspect of likelihood, and consequences and calculate risk levels. All three mentioned criteria must be set in such a way that the analysis observes and projects in real-time with a large knowledge base that is the basis for calculating parameter values.

A risk evaluation must provide a clear answer to a question that is not just a simple number but a multidimensional framework that shows trends and every single evaluation parameter in real-time.

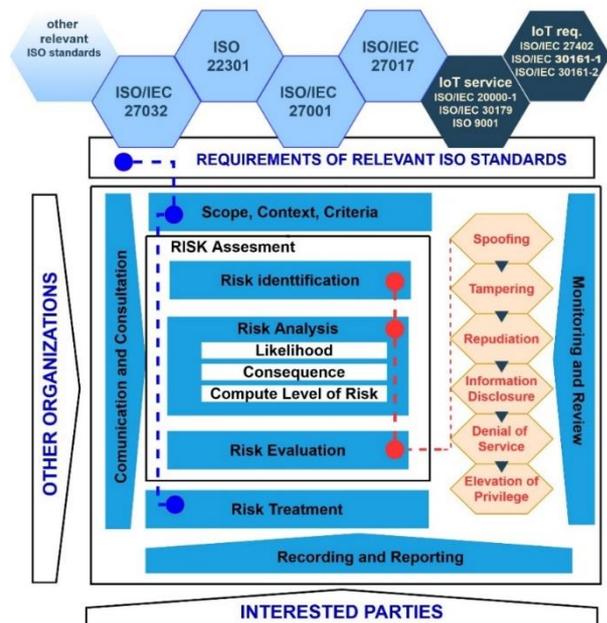


Figure 6: RISK assessment (based on ISO 27005) extended with STRIDE cycle and ISO requirements

STRIDE is a threat modeling methodology used to identify potential security threats in a system and develop appropriate countermeasures. It is an acronym for six types of threats: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege. STRIDE threat modeling is connected to the stages of Risk assessment.

Risk treatment as the next phase is mainly oriented towards acting according to each individually identified requirement of the ISO standard within the matrix. The knowledge acquired in the process of treating risks during the management of the information security system is a significant source for further analysis using the principles of machine learning and artificial intelligence.

STRIDE, threat modeling methodology cannot be established by simply adopting a good methodology, but it needs to be placed alongside existing risk treatment methodologies and in a specific environment that is analyzed by defining the requirements by adopting the requirements of certain specific standards for each specific environment. There is no magic wand that will solve the problems of information security, On the contrary, a complex, multi-layered, and multi-dimensional framework must be established that will require a complete and detailed analysis of the system using a process approach and a complete analysis of threats, vulnerabilities, and risks. The mentioned activity is not a one-time activity but represents a cyclical repetition of the identified elements in each subsequent time parameter that is partially (and in some elements completely) different from the previous one. There are two basic misconceptions that need to be removed from the thinking of the multidisciplinary team: the input parameters for the risk treatment are immutable, and the risk treatment model is immutable. The presented model RISK assessment (based on *ISO 27005*) extended with STRIDE cycle and ISO requirements is an example of a basic layer developed on many years of experience in the field of information security.

5.1. Identification of ISO requirements for information security and IoT

The initial framework for action in the domain of information security is defined within the requirements of the following standards:

- *ISO/IEC 27001*: Information security, cybersecurity and privacy protection - Information security management systems - Requirements,
- *ISO/IEC 27701*: Security techniques - Extension to *ISO/IEC 27001* and *ISO /IEC 27002* for privacy information management - Requirements and guidelines,

- *ISO/IEC 27011*: Information security, cybersecurity and privacy protection - Information security controls based on *ISO/IEC 27002* for telecommunications organizations,
- *ISO/IEC 27032*: Cybersecurity - Guidelines for Internet security,
- *ISO/IEC 22301*: Security and resilience - Business continuity management systems - Requirements and
- *ISO/IEC 20000-1*: information technology - Service management - Service management system requirements

Within enlisted requirements of the standard, it is certainly and inevitably necessary to treat the requirements of the *ISO 9001*: Quality management systems - Requirements standard in the domain of detailed and complete analysis of opportunities and risks as part of a complete and detailed analysis of the organization's processes.

A detailed risk analysis of the process is the place where threats, vulnerabilities, and risks must be fully defined, with a clear and complete connection to the parameters of the information system. It is especially important to respect the dynamism of changes in the initially identified parameters, connections, importance, and influence within the previously defined frameworks. A complete analysis of the process through the implementation of *ISO 9001* requirements is the first and basic phase of the successful design of an information system that bases information security on risk management [10].

From this initial phase, must be established a clear risk matrix that is related to specific process entities and risks at the base level. The implementation of the requirements of the *ISO/IEC 27032* standard through the formation of a Cyber Security Program must be carried out for the entire organization, including all components of information systems through business and IT functions, given the fact that attacks and threats to the information security system can appear from every side and in every moment. It is necessary to implement the requirements of the standards *ISO/IEC 27001*, *ISO/IEC 27701*, *ISO/IEC 27032*, *ISO/IEC 20000-1*, and *ISO/IEC 22301* in areas that include security in the intranet/internet space, i.e. intranet/internet security issues that focus on bridging the results of risk analysis between different domains of information security in the intranet/internet space.

According to the stated standards, it is necessary to implement technical guidelines for solving intranet/internet security risks, including social engineering attacks, hacking, spyware, and attacks using potentially malicious malware software. These technical guidelines should provide controls for the treatment of all identified risks, including controls for preparing responses to attacks from malicious software (malware), reactions to detected unusual events related to information security and monitoring the attack architecture. It is necessary to create a framework for efficient and effective information exchange, coordination, and incident management among related stakeholders in the predictively defined cyberspace of the organization. Interested parties that may be involved are employees, customers, and third parties, which may be different types of organizations or individuals, as well as providers, which include service providers as well as all those identified by the risk matrix.

The analysis of interested parties should not be limited only to cyberspace but should perform a complete analysis of the technological, procedural, and organizational framework, not neglecting the social and other specific general security aspects of the organization.

The corrective reaction is not a satisfactory achievement. The preventive reaction is satisfactory only in the process of implementation of the system. Predictive response is a target function that must be established in the system.

On the example of the identification of threats, vulnerabilities, and risks within the requirements of the *ISO/IEC 27001* standard, the same methodological basis should be established for the requirements of the *ISO/IEC 27701*, *ISO/IEC 27032*, and *ISO/IEC 27011* standards, as well as for the requirements of the *ISO/IEC 20000-1* and *ISO 22301*.

Special attention should be directed to the identification of IoT needs. The initial framework for IoT identification is covered by the following standards:

- *ISO/IEC 27402*: Cybersecurity - IoT security and privacy - Device baseline requirements,
- *ISO/IEC 30161-1*: Internet of Things (IoT) - Data exchange platform for IoT services - Part 1: General requirements and architecture,
- *ISO/IEC 30161-2*: Internet of Things (IoT) - Data exchange platform for IoT services Part 2: Transport interoperability between nodal points,

- *ISO/IEC 30179:2023*: Internet of Things (IoT) - Overview and general requirements of IoT system for ecological environment monitoring, but also other specifically identified for individual system.

6. Conclusion

The framework for the implementation of ISO ICT security standards in IoT systems is presented in this paper. Along with the set of standards that should be implemented to achieve security in IoT environments, the recommendations for their implementation are given. Also, the usage of threat modeling tools for modeling IoT systems and identifying the threats for the modeled systems is presented. The reports of the IoT systems threat modeling are used as input values for a given framework. The proposed methodology is illustrated on the example of a simple IoT system designed for urban pollution, and noise monitoring.

7. References

- [1] V. Drake, OWASP, Threat Modeling, Retrieved May 2024. URL: https://owasp.org/www-community/Threat_Modeling.
- [2] A. Konev, A. Shelupanov, M. Kataev, V. Ageeva, A. Nabieva, A Survey on Threat-Modeling Techniques: Protected Objects and Classification of Threats. *Symmetry*. 2022; Vol. 14(3):549. <https://doi.org/10.3390/sym14030549>
- [3] Y. C. Tok, S. Chattopadhyay, Identifying threats, cybercrime and digital forensic opportunities in Smart City Infrastructure via threat modeling, *Forensic Science International: Digital Investigation*, Vol. 45, 2023, 301540, <https://doi.org/10.1016/j.fsidi.2023.301540>.
- [4] V. Casola, A. De Benedictis, M. Rak, U. Villano, Toward the automation of threat modeling and risk assessment in IoT systems, *Internet of Things*, Vol. 7, 2019, 100056, <https://doi.org/10.1016/j.iot.2019.100056>.
- [5] A. Gerodimos, L. Maglaras, M. A. Ferrag, N. Ayres, I. Kantzavelou, IoT: Communication protocols and security threats, *Internet of Things and Cyber-Physical Systems*, Vol. 3, 2023, pp 1-13, <https://doi.org/10.1016/j.iotcps.2022.12.003>.
- [6] S. Sicari, A. Rizzardi, D. Miorandi, A. Coen-Portisini, A risk assessment methodology for the Internet of Things, *Computer Communications*, Vol. 129, 2018, pp 67-79, <https://doi.org/10.1016/j.comcom.2018.07.024>.
- [7] A. Y. Wong, E. G. Chekole, M. Ochoa, J. Zhou, On the Security of Containers: Threat Modeling, Attack Analysis, and Mitigation Strategies, *Computers & Security*, Vol. 128, 2023, 103140, <https://doi.org/10.1016/j.cose.2023.103140>.
- [8] P. Sethi and S. R. Sarangi, "Internet of Things: Architectures, Protocols, and Applications", *Journal of Electrical and Computer Engineering*, Vol. 2017, 25 pages, <https://doi.org/10.1155/2017/9324035>
- [9] D. Dobrilović, V. Brtko, G. Jotanović, et al. The urban traffic noise monitoring system based on LoRaWAN technology. *Wireless Networks*, Vol. 28, pp 441–458, 2022, <https://doi.org/10.1007/s11276-021-02586-2>
- [10] R. Dragović, D. Dobrilovic, D. Dragović (2022, September). Recommendations for the Creation of Usable Critical Infrastructure for the Delivery of Priority Services of State Bodies. In *IFIP International Conference on Human Choice and Computers* (pp. 161-171). Cham: Springer Nature Switzerland. DOI:10.1007/978-3-031-47990-8_15

Services for Web-Publishing and Extraction of Fragments of Big Raster Data-sets

Alexei E. Hmelnov¹

¹*Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences (IDSTU SB RAS), 134 Lermontov St, Irkutsk, 664033, Russian Federation*

Abstract

To be able to conveniently work with large raster data-sets we have developed the MRG file format. In the article we consider the software tools developed to create and process the MRG files and the global DEMs and the other global raster data that we have collected and converted to MRG. To make the data available for non local users we have developed the WMS and WPS services for publishing MRG data. In the article we discuss some details of their implementation and usage.

Keywords

data format, big raster data, loss-less compression, global DEM, WMS service, WPS service

1. Introduction

Since the time, when in the beginning of 2000th the 1st version of Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) had been published, the progress of remote sensing technologies has led to public availability of several global digital elevation models (GDEM) and now we can choose which of them to use for a particular project. The quality of the models may strongly vary, for example, while the ASTER GDEM is usually reported to be more precise than SRTM, our experience shows that in the area of Irkutsk it has prohibitively low quality, so that we don't even consider this GDEM in our current research.

The best resolution of the current publicly available GDEMs is 1'' (arc-second) $\approx 30m$ per pixel. All the GDEMs are published as a collection of tiles with the sizes from $1^\circ \times 1^\circ$ to $5^\circ \times 5^\circ$ in GeoTIFF or JPEG 2000 file formats. So all the tiles of typical GDEM take several dozens of gigabytes. To obtain the tiles it is usually required to register on the corresponding site, and it will take rather long time to find and download the data for an area of interest, and the user sometimes may be banned for trying to download the data too fast. Because the end-user is usually interested in exploring a particular area of GDEM, after downloading the data it will usually be required to combine the tiles into a single raster file and to cut the area of interest from it. As a result it takes quite a lot of time to start the actual usage of the data to solve the problem under consideration.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ hmelnov@icc.ru (A. E. Hmelnov)

🆔 0000-0002-0125-1130 (A. E. Hmelnov)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (iccs-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.07

After we found that the SRTM data become publicly available we have developed the file format MRG (MultiResolution Grid), which is intended for compact storage of big raster images with the capability to quickly extract any fragments of the image with any resolution [1]. When represented in MRG, DEM data usually take less space, than the original tiles, even compressed, but, in contrast to the original formats, in MRG we have the random access to any DEM fragments.

In contrast to the MrSID, ECW or JPEG 2000 file formats, which also have some kind of resolution pyramids, but use very complex and time consuming transforms, the MRG file format is based upon construction of a resolution pyramid and upon loss-less compression of the differences between the interpolated and the actual data for the next resolution level using a family of very simple compression algorithms. The complexity of these algorithms is always asymmetric: while the compression phase searches for the optimal split of the sequence of delta values into the intervals of constant bit depths using the dynamic programming method, the decompression phase is extremely fast, because it simply reads and decodes the intervals [2].

Representation of a DEM in MRG makes it seamless and gives the user the capability to forget about the tiles and to quickly pan and zoom into any DEM part and to extract any its fragment with the required level of details. So, the local user can quickly find the data of interest and start working with it.

Recently with the support of the big project “Fundamentals, methods and technologies for digital monitoring and forecasting of the environmental situation on the Baikal natural territory” we have acquired a data storage system of the size of several hundred terabytes. Using the data storage it becomes possible to collect several GDEMs and to convert them into the corresponding MRG files. We had to advance somehow the MRG software to be able to work with the data as big as never before.

To provide convenient access to the big MRG data-sets for the remote users we have developed a couple of services: the WMS service for viewing the data, and the WPS service for extraction of their fragments.

In the article we'll consider some details of the implementation of the services and their usage and some statistics on the examples of the GDEM and optical remote sensing data represented in MRG.

2. The MRG software

To work with the MRG files we have developed a number of software modules that allow us to create the files from the original GeoTIFF, JPEG 2000 and the other file formats, view the resulting raster images, and retrieve from the MRG files various kinds of information, like image fragments with desired resolution, contour lines or cross-sections:

MRGView is an interactive (GUI) program for processing DEMs;

MRGImgV is an interactive (GUI) program for processing multichannel images;

MRGLib is a dynamic library, that makes it possible to develop external programs, which work with MRG;

MRGCmd is a command line utility, that can perform in batch mode almost all the operations, that we can do in MRGView and MRGImgV;

MRGWebSrv is a standalone application that works as a Web-server and implements the WMS-service and WPS-service protocols for publishing the registered MRG files;

MRGWebSrvSvc is a Windows service that works the same way as MRGWebSrv.

All the MRG software use the common code base. We use MRGWebSrv mostly for debugging purposes, because it is much more easier to debug a standalone application, than the Windows service. And the main advantage of Windows service is that after its registration it works in an unattended mode and will be automatically restarted, when the computer restarts.

The MRGCmd application plays crucial role for construction of big GDEM files, because, in contrast to the GUI applications, it allows us e.g. to rerun the same task several times with exactly the same parameters, when something goes wrong, or to change some parameters and process exactly the same data again to compare the results of the two runs. The MRGCmd utility can perform a sequence of the following commands

- **Set** - Set MRG creation options
- **Open** - Open MRG file
- **NewImage** - Create MRG file from images
- **NewDEM** - Create MRG file from DEM files
- **Describe** - Describe MRG
- **GetIsolines** - Get isolines from MRG image
- **GetProfiles** - Get profiles from MRG image
- **Export** - Export fragment to GeoTIFF
- **Defragment** - Create defragmented version of MRG
- **Compare** - Compare current MRG to another one
- **Exec** - Execute commands from file

which can be passed in the application command line separated by semicolons or in a separate text file using the Exec command.

Let's consider a couple of examples. The following command compares two files: the old and already checked file testWok.mrg and the new file testW.mrg. The comparison will stop after finding 4 errors (mismatches).

```
MRGCmd.exe Open testWok.mrg ; Compare -e4 testW.mrg
```

And the following bat-file generates tif previews for a series of MRG files, that were created for the Copernicus DEM bands of different horizontal resolution.

```
for %%f in (all_*.mrg) do (  
    echo %%f  
    MRGCmd Set -W ; Open %%f ; Export -f -15 TIFF\%%~nf.tif  
)
```

The -15 flag tells the Export command to use the 5th level of details, i.e. to use the resolution which is $2^5 = 32$ times lower than the original very large image.

3. The Global DEMs and the other raster images represented by MRG files

As it was already mentioned above, the very development of the MRG file format was inspired by the open publishing of STRM data. The 1st version of SRTM data-set had resolution of $3'' \approx 90m$ per pixel, and it takes approximately from 8 to 10 GB to represent the data in MRG file format (the difference in the file sizes is explained by the choice of the interval header packing algorithm: the default one gives about 10 GB, and the algorithm based upon the interval statistics for these data gives better results, but to obtain the statistics it is anyway required first to compress the data with the default parameters).

Initially we were not trying to create the GDEMs, and were completely satisfied by the fact that in the MRG file it was possible to store the whole DEM for the territory of Russia on a single CD and to quickly view the image immediately from this CD. When creating the MRG file for the whole SRTM GDEM it was first required to upgrade the MRG file format, because the 32-bit file offsets of the data blocks, which were initially used in the files, had become no longer enough for the 10 GB files. To solve the problem we have added to the file header a field about the data block offset alignment: now the default alignment of 16 bytes enabled us to create the files with the size up to 64GB. But to represent the modern global DEMs and the other raster images this limitation may become too restrictive. Increasing the block alignment further would mean wasting much more space, so we decided to switch to the 64-bit offsets for the very large files of this kind. The modern SRTM GDEM of the $1''$ per pixel resolution [3] takes up 56.3 GB in MRG (Fig. 1), while the source data were published in 14 297 HGT files (pure 2D arrays of 16-bit integer values) compressed by ZIP of the total size 97.6 GB.

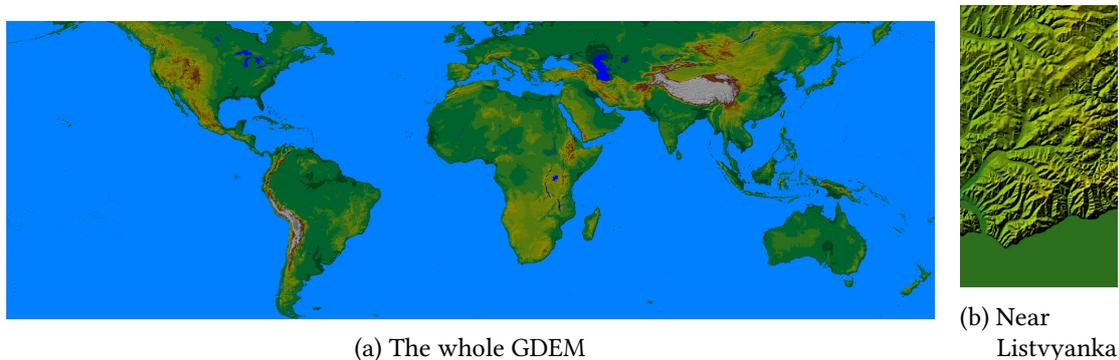


Figure 1: The hill shading of SRTM GDEM from MRG

According to some modern research papers [4], which compare different GDEMs, the best of them, which is far ahead of the others, is the Copernicus DEM. In fact the Copernicus DEM was produced from WorldDEM – a commercial GDEM with the initial resolution of 10 m per pixel, that was then improved in its next version WorldDEM Neo to 5 m per pixel. So, while, say, the SRTM DEM is a result of the attempts to up-sample the initial DEM of lower resolution using every trick in the book, the Copernicus DEM is a result of down-sampling of the much more precise WorldDEM data.

The Copernicus DEM data for 2022 as it was published [5],[6] take 1.74 TB and contain 26 450 TAR files. Each of the uncompressed TAR archives contains single DEM tile in an uncompressed GeoTIFF file and the other uncompressed auxiliary, preview, and license files. Among the auxiliary files the largest is the Height Error Mask GeoTIFF, which always has exactly the same file size as the DEM GeoTIFF, because it is also uncompressed and has the same pixel size and format. As a result of adding all the auxiliary files the DEM data take up about 47% of the TAR file size. We believe that it would be much easier to download and distribute the data if the tiles were compressed somehow.

The pixel values of the Copernicus DEM GeoTIFFs are the 32-bit floats. Unfortunately, the compression algorithm in the MRG file format is designed for integer differences only, and by now it can contain only the 16-bit integer values, so we have to round the float values to 0.25 m steps to be able to fit the height of Qomolangma into the unsigned 16-bit range.

Another peculiarity of the Copernicus DEM data is that the horizontal resolution of the tiles varies with latitude. All the tiles cover the area $1^\circ \times 1^\circ$, but their horizontal (by longitude) resolution varies from the same as the vertical (by latitude) resolution of 3600 pixels per degree in the stripe of latitudes from -50° to 49° to 360 pixels per degree for the Antarctic tiles with the latitudes below or equal to -86° (see Table 1). The latitudes in the names of the tiles are the latitudes of their lower (south) edges, so the north and south stripes are in fact symmetric (mirror each other). As a result we have to split the data into 10 stripes of the consecutive tiles having the same horizontal resolution. The total size of the Copernicus DEM data in MRG is 96.5 GB.

Table 1: Horizontal resolution of the Copernicus DEM tiles

Latitude	Pixels/°	MRG size
$-90^\circ..-86^\circ$	360	0,98 GB
$-85^\circ..-81^\circ$	720	2.05 GB
$-80^\circ..-71^\circ$	1200	5.25 GB
$-70^\circ..-61^\circ$	1800	1.20 GB
$-60^\circ..-51^\circ$	2400	151 MB
$-50^\circ..49^\circ$	3600	66.7 GB
$50^\circ..59^\circ$	2400	9.64 GB
$60^\circ..69^\circ$	1800	8.26 GB
$70^\circ..79^\circ$	1200	2.02 GB
$80^\circ..84^\circ$	720	267 MB

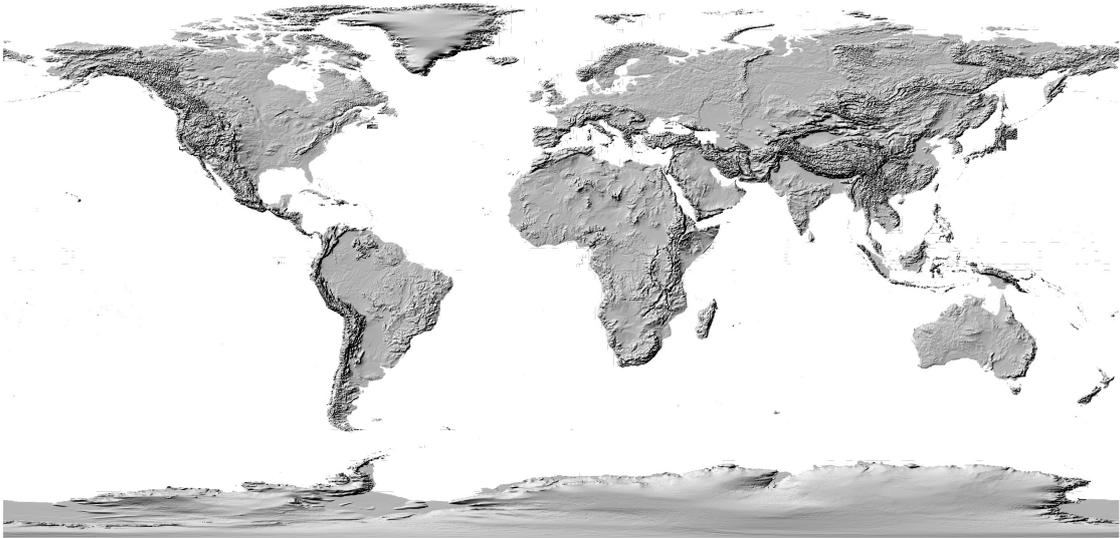


Figure 2: The shadow relief of Copernicus DEM in 10 MRG files shown together

Another interesting GDEM, that was produced from the Copernicus DEM using machine learning is FABDEM (Forest And Buildings removed Copernicus DEM) [7],[8]. This data-set is a result of an attempt to make some approximation of global DTM from Copernicus DEM by subtracting the heights of trees in the area of forests and the heights of buildings in the areas of cities, which are estimated using the random forest algorithm trained on a series of sample areas, for which the heights were measured by the means of more precise instruments. By visual inspection of the FABDEM we have found some artifacts in the data, like the numerous stepped boundaries which are clearly visible on the Figure 3(b) (which indicate us that the FABDEM generation algorithm usually ‘deforest’ the Copernicus DEM using some rectangular areas), but the overall quality of the data-set is reported to be very high.

The whole FABDEM data-set is distributed in 356 zip archives, each of them covers an area of $10^\circ \times 10^\circ$ and may contain up to 100 files for the $1^\circ \times 1^\circ$ tiles in the area. The pixel values of the FABDEM GeoTIFFs are also the 32-bit floats. But in contrast to the Copernicus DEM data the horizontal resolution of all the FABDEM tiles was risen to 3600 pixels per degree. The total size of the FABDEM archives is 296 GB, and the size of the resulting MRG file with the default header compression parameters is 85.3 GB, but here we also have rounded its values to the same 0.25 m steps.

Another useful global data-set is the high-resolution canopy height model of the Earth (the Global Canopy) [9], [10]. The vegetation height model has been obtained from the Sentinel images using an ensemble of convolutional neural networks (CNNs) and has the corresponding global resolution of 10 m/pixel, which is 3 times higher than all the models considered herein-before. The Global Canopy data are published in 2650 GeoTIFF files with 8-bit integer pixel values, representing the estimated tree heights in meters. Almost all the heights are in the range from 0 to 45 m and in the areas without vegetation like water objects or deserts the heights has the ‘no data’ values. The GeoTIFF files use the LZW compression and their total size is 349 GB. The dimensions of the resulting MRG file (Figure 4) are 4320000×1728000 pixels, and its size is 331 GB when using the default interval header compression algorithm, which in fact

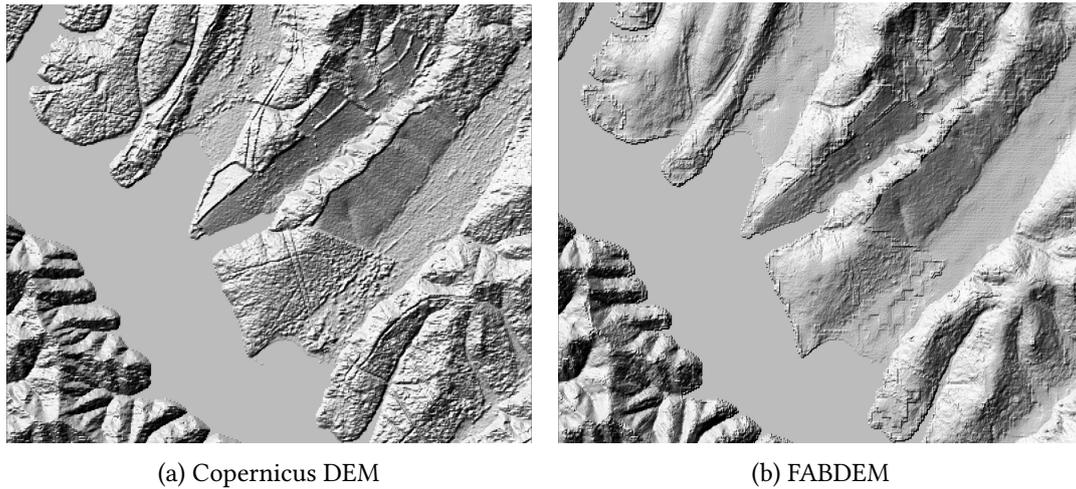


Figure 3: Copernicus DEM vs FABDEM (the hill shading of the same area)

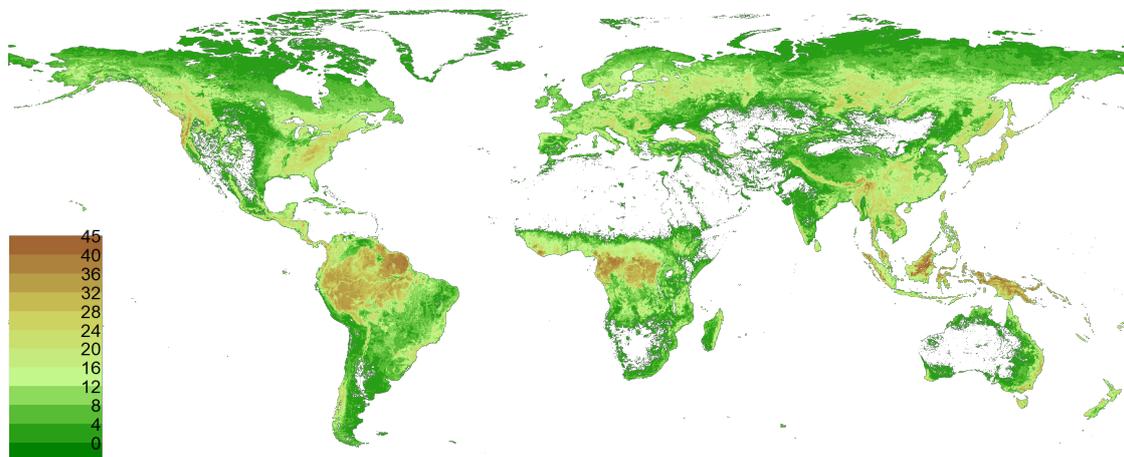


Figure 4: The Global Canopy data-set. Its pixel values are the estimated tree heights in meters

is more suitable for the 16-bit values. After using a more specific to the data interval header packing, which is based upon the statistics of the data, the resulting MRG file size further goes down to 323 GB.

The last but not least in our collection of the global data-sets is a by-product of the project known as the Hansen forest [11][12]. The main goal of the project is to observe the global tree cover changes using the Landsat images. As a result of the project its authors have produced such data-sets as ‘Forest loss year’ or ‘Tree canopy cover for year 2000’ but to obtain the data they have also produced the global composite images for a series of years including the last of them ‘Circa year 2022 Landsat cloud-free image composite’. This composite image has its own independent value which is of interest for us: it can be used as a base layer for various maps. Unfortunately for us the composite image doesn’t include all the true-color channels, it contains

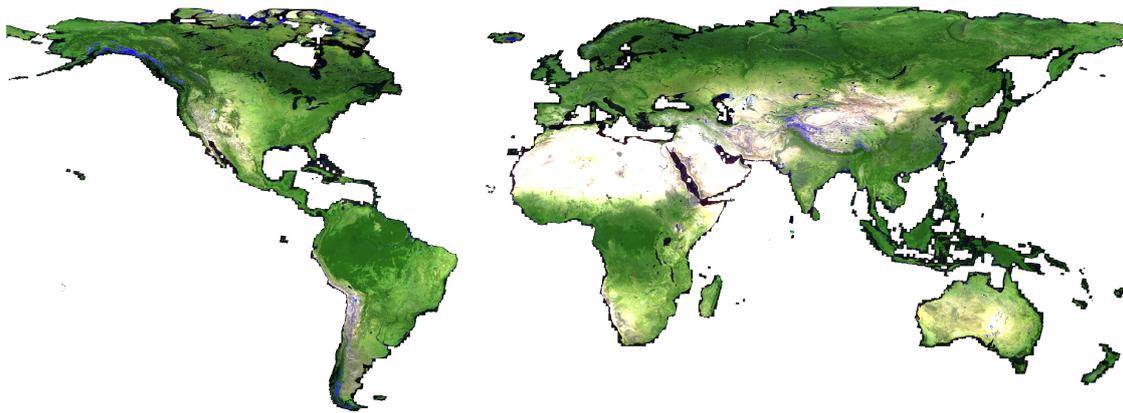


Figure 5: The Hansen forest circa year 2022 Landsat cloud-free image composite in pseudo colors: red=SWIR2, green=SWIR1, blue=Red

only the following 4 Landsat bands, which are needed for the forest analysis: Red ($0.66 \mu m$), NIR ($0.86 \mu m$), SWIR1 ($1.6 \mu m$), SWIR2 ($2.2 \mu m$), but some combinations of the bands may look rather natural (Figure 5).

In contrast to the other MRG files the Hansen forest composite image has more than one (four) channels, and it uses the color decorrelation algorithm to improve the compression ratio [13]. The global Hansen forest composite image data are published as 504 GeoTIFF tiles with LZW compression, which have four 8-bit channels (Red, NIR, SWIR1, SWIR2). The total size of the files is 755 GB, and the size of the corresponding MRG file is 446 GB.

4. Web service for publishing MRG files

It would be impossible to immediately use the raster images of the sizes as it was described in the previous chapter in any modern GIS. So we have developed the program `MRGWebSvc`, which implements two OGC standards for accessing the registered big MRG files: WMS and WPS. Initially we have supported the WMS (Web Map Service) standard, that allows users to view the images as a map layers and combine the layers with any other raster or vector data-sets. To be published by the service the MRG files should be registered in its configuration file. Besides from the MRG file path the configuration should specify its display name and display parameters.

To be able to work with the fragments of big MRG files in GIS programs we have additionally implemented the WPS (Web Processing Service) processes for requesting the data. To obtain the the list of processes of a WPS service the `GetCapabilities` request can be used. And the `DescribeProcess` request provides the information about the inputs and outputs of the process. To run the process the `Execute` request is used. The WPS standard imposes certain restrictions on the ways of interaction with the service, which do not always make this interaction as convenient as it may be without these restrictions: the service should implement some *processes*, each process has a number of *input* and *output* parameters; the parameters can be passed by value (the *literals*) or in the form of a fragment of the XML code of request

or response or by URL (for *complex data*); if a process has single output parameter, it can immediately return its value in the response to the `Execute` request.

There exist software libraries for implementation of WPS services, for example, the 52°North WPS library for Java or PyWPS or OWSLib for Python, but in fact for a small service it may be easier to implement the processing of the WPS requests from scratch.

In the current version of the WPS service for obtaining raster fragments we have implemented two processes: `Export`, which returns the fragments in GeoTIFF file format, and `QueryPoints`, which retrieves the DEM values in the list of points. The `Export` process has the following input parameters: `name` – the name of a registered raster; `bboxin` – bounding rectangle of the exported area; `level` – the level of details (0 denotes the most detailed level, corresponding to the resolution of the MRG file, then increasing the level by 1 reduces the resolution by half). It returns GeoTIFF file with the contents of the specified fragment of the selected raster at the specified level of details.

The value of the `name` parameter should belong to the list of names of the registered MRG files. The WPS standard allows us to specify the list of valid values of the input literal `name` in the `DescribeProcess` response. The WPS client, which can use this information, may create a drop-down list for the name selection field. This method turns out to be not very convenient when there are a lot of registered MRG files: while the WMS protocol allows us to organize the map layers into a tree, here we are limited to a linear list only. And the more serious limitation of the WPS protocol is that we can't specify the valid range of values of the `bboxin` and `level` parameters in the process description, since these constraints depend on the selected layer, and this kind of inter-dependencies is not supported by the WMS standard. It would be possible to implement additional processes for obtaining this kind of meta-information about the registered MRG files, but the WPS standard still will not support the automated checking of these constraints.

To get the image of a MRG file with the minimal level of details and simultaneously find out which area it covers, one can pass the `level = -1` (here we use the negative indexes in the style of the Python arrays) and the default `bboxin`, which corresponds to the maximum bounding box ($-180^\circ -90^\circ; 180^\circ 90^\circ$).

The way the resulting image fragment will be returned by the `Export` process depends on its size, the size limits can be set in the service configuration file. The following threshold values are currently selected: the images with the size of up to 4Mpx are directly returned; if the number of pixels exceeds this threshold, the file is first saved in a special folder, which can be accessed by the geoportal Web server, and the `Execute` request returns the XML document with the URL corresponding to this file on the Web-server. Since the MRG format's capability to store large-scale raster data is significantly superior to that of GeoTIFF, it makes no sense to try to get all the MRG information at level 0 in one request: even if the file were created, it would be impossible to work with it. Therefore we have another threshold: the maximum image fragment size. The current value of this threshold is 1024Mpx.

For testing WPS services in QGIS, we were able to find only one plug-in – the WPS Client which belongs to the experimental category. However, the capabilities of this plug-in turned out to be sufficient for debugging our service. In particular, it creates a drop-down list of valid values for the `name` parameter, and also it automatically substitutes the bounding box of the current map view for the value of the `bboxin` parameter (Figure 6(a)). And to visually select

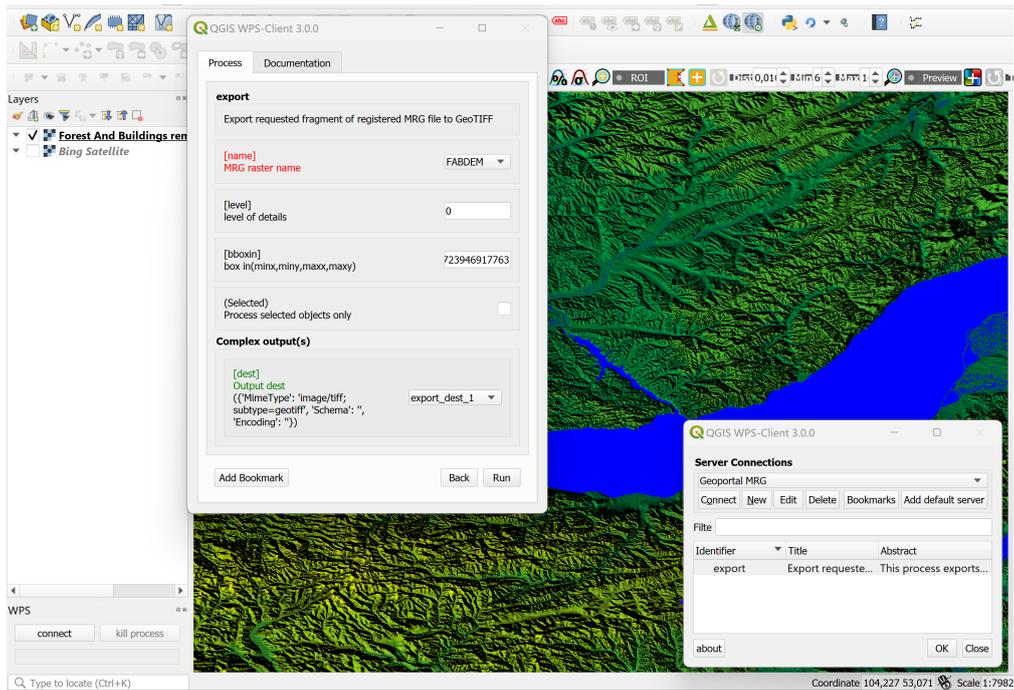
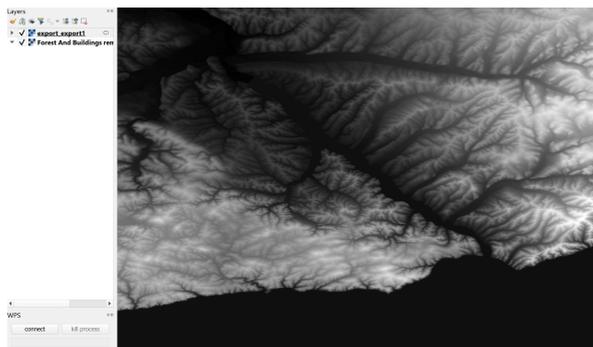


Figure 6: Request of MRG file fragment using the WPS Client plug-in dialog in QGIS

the area of interest we can use the images obtained from the WMS service of our MRG server.

After executing the request, a temporary raster layer is created in QGIS, which can be further used to solve various problems of processing and analyzing the received data. For example, different terrain models can be compared to each other (Figure 6(b)). And the use in QGIS is not the main purpose of the developed WPS service: the geo-portal software allows its users to create compositions of services, which can now include the data fetched from the MRG format.



(a) New layer added as a result of the WPS request



(b) Comparison of Copernicus DEM and SRTM in QGIS

Figure 7: Processing the requested MRG file fragments in QGIS

5. Conclusion

The MRG file format and the software developed for its processing can substantially streamline access to the global raster data-sets, that are available now. The data-sets in MRG format take less space than in their original representations, and the main advantage of this representation is that any data fragment can be quickly accessed with any level of details.

The services developed for publishing MRG data follow the OGC WMS and WPS standards and allow non local users to easily access the data. The resulting data layers make their contribution into our geo-portal data collection. Moreover, the WPS service results can be used for development of more complex data processing algorithms. For example, the information about point elevations from the GDEMs may be used for training neural networks.

References

- [1] A. E. Hmelnov, The MRG file format for compact representation and fast decompression of large digital elevation models (in russian), *Computational technologies* 20 (2015) 63–74.
- [2] A. E. Hmelnov, A lossless compression algorithm for integer differences sequences by optimization of their division into intervals of constant bit depth values (in russian), *Computational technologies* 20 (2015) 75–98.
- [3] Shuttle radar topography mission (SRTM) global, 2013. URL: <https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model>. doi:10.5069/G9445JDF.
- [4] C. Bielski, C. Lopez-Vazquez, C. H. Grohmann, P. L. Guth, L. Hawker, D. Gesch, S. Trevisani, V. Herrera-Cruz, S. Riazanoff, A. Corseaux, H. I. Reuter, P. Strobl, Novel approach for ranking DEMs: Copernicus DEM improves one arc second open global topography, *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024) 1–22. doi:10.1109/TGRS.2024.3368015.
- [5] Copernicus DEM - global and european digital elevation model (COP-DEM), 2022. URL: <https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model>. doi:10.5270/ESA-c5d3d65.
- [6] M. Buchhorn, M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, B. Smets, Copernicus global land cover layers–collection 2, *Remote Sensing* 12 (2020). URL: <https://www.mdpi.com/2072-4292/12/6/1044>. doi:10.3390/rs12061044.
- [7] L. Hawker, P. Uhe, L. Paulo, J. Sosa, J. Savage, C. Sampson, J. Neal, A 30 m global map of elevation with forests and buildings removed, *Environmental Research Letters* 17 (2022) 024016. doi:10.1088/1748-9326/ac4d4f.
- [8] J. Neal, L. Hawker, FABDEM V1-2, 2023. URL: <https://data.bris.ac.uk/data/dataset/s5hqmjcdj8yo2ibzi9b4ew3sn>. doi:10.5523/bris.s5hqmjcdj8yo2ibzi9b4ew3sn.
- [9] N. Lang, W. Jetz, K. Schindler, J. D. Wegner, A high-resolution canopy height model of the Earth, *Nature Ecology & Evolution* 7 (2023) 1778–1789. URL: <https://doi.org/10.1038/s41559-023-02206-6>. doi:10.1038/s41559-023-02206-6.
- [10] N. Lang, W. Jetz, K. Schindler, J. D. Wegner, A high-resolution canopy height model of the Earth, 2023. URL: <https://langnico.github.io/globalcanopyheight/>.

- [11] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, J. R. G. Townshend, High-resolution global maps of 21st-century forest cover change, *Science* 342 (2013) 850–853. URL: <https://www.science.org/doi/abs/10.1126/science.1244693>. doi:10.1126/science.1244693. arXiv:<https://www.science.org/doi/pdf/10.1126/science.1244693>.
- [12] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, J. R. G. Townshend, Global forest change, 2023. URL: <https://glad.earthengine.app/view/global-forest-change>.
- [13] A. E. Hmelnov, Reversible integer approximation of color space transforms for lossless compression of big color raster data, *Computer Optics* 46 (2022) 492–505. doi:10.18287/2412-6179-CO-1052.

Comparison of Table Type Taxonomies

A. Shigarov^{1,2}, V. Paramonov^{1,2}

¹Matrosov Institute for System Dynamics and Control Theory, Siberian Branch of Russian Academy of Sciences, 134 Lermontov str., Irkutsk, Russia

²Institute of Mathematics and Information Technologies, Irkutsk State University, 20 Gagarina ave., Irkutsk, Russia

Abstract

Tables are widely used to present related data in various formats and media. In particular, an immense number of HTML tables can be found on web pages. These tables serve as a valuable source of data for applications such as web mining, question-answering systems, and knowledge base development. However, not all HTML tables contain actual relational data. Many of them are used primarily for layout and navigation purposes. On the other hand, genuine tables may have different layouts, formatting, and content. One of the challenges in extracting data from web tables is to identify the main functional and layout types and properties of the tables that are commonly found on the internet. Currently, we have a wide range of taxonomies that classify table types and their properties. These taxonomies can be used to categorize tables and select the appropriate method for processing tabular data based on their specific type. Although the existing taxonomies provide a variety of table types, they often use confusing terminology. This paper aims to provide an overview of these taxonomies by aligning their terminology and comparing them in a qualitative manner.

Keywords

table understanding, table extraction, type classification, table layout

1. Introduction

Tables are a one way to organise and present multiple sets data efficiently. They are widely used on the web and often embedded in HTML pages. These tables can contain various types of data, including unstructured, semi-structured, and relational data. Approachers to automated tables processing depend on their type and the way of data organisation.

According to one of the first studies on table classification, presented by Wang et al. [1], all tables are divided into two categories: *genuine* and *non-genuine*. A *genuine* table includes relational data, where the table structure helps convey the relationships between the data in the cells. In contrast, *non-genuine* tables are used for layout purposes.

Extracting, integrating, and analysing information from *genuine* tables can help uncover new knowledges [2, 3]. All genuine tables can be classified into different types based on how the data is organized. Understanding these types can improve the quality of data processing. Currently, there are some studies discussing table taxonomies [4, 5, 6, 7, 8, 9, 10, 11, 12]. Exploring these taxonomies can lead to better data extraction from HTML tables and understanding them.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ shigarov@icc.ru (A. Shigarov); slv@icc.ru (V. Paramonov)

ORCID 0000-0001-5572-5349 (A. Shigarov); 0000-0002-4662-3612 (V. Paramonov)

 © 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 ICCS-DE 2024 Workshop Proceedings (icc-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.08

HTML tables are commonly classified on their layout. Defining the type of a table can simplify data extraction and analysis [1, 13, 4]. In this analysis, we aim to establish which cells contain meta-data or semantic hints that help understand the meaning of the data, and which cells hold actual data. Further, in all figures 1, 2, 3, 4, 5, 6 cells with meta-data have a blue background, and cells with data have a white background.

This paper focuses on analysing existing taxonomies for web tables in HTML. The main goal is to identify relationships between these taxonomies and understand how they can be applied to improve data extraction and analysis from HTML tables.

2. Existing Table Type Taxonomies

In the study [13] the goal of Cafarella et al. research is to extract a set of high-quality relations [13] from HTML documents. In the research, authors present the fundamental division of whole tables into two types: *relational* and *non-relational*, which in the other classification [1] are designated as *genuine* and *non-genuine*. The results of Crestan’s research, which became a basis of further research, presented in the paper [4]. During the analysis of these tables, a broad classification and set of problems in extracting tables of certain types were established. This study divides the entire set of tables into *relational knowledge* tables and *layout* tables, which are equivalent to *genuine* and *non-genuine* accordingly in [1]. *Layout tables* are further divided into *navigational* and *formatting*. However, these table types are not relevant in our study. Below we consider different types of tables according to the Crestan’s taxonomy [4] and give examples to illustrate table reliability to a certain type.

- *Vertical listings*. A table that lists the attributes of several entities that are similar in meaning (Fig. 1, a).
- *Horizontal listings*. Tables are similar to a *vertical listings*, but with a different orientation. This type of tables often is used for comparison (Fig. 1, b).
- *Attribute/value*. A special case of vertical and *horizontal listings*, that does not contain a subject (Fig. 1, c).
- *Matrix*. Tables that have values at the junctions of rows and columns, while having headers on the left and top of the table (Fig. 1, d).
- *Calendar*. A special case of *matrix tables* in which the subject is a date and the predicate is some kind of event (Fig. 1, e).
- *Enumeration*. The tables list a number of objects that have a similar meaning (Fig. 1, f).
- *Form*. Tables consist of input fields that the user can fill in or select (Fig. 1, g).
- *Other*. This type includes tables that cannot be attributed to the types listed above.

In 2013, Lautert [5] presents a classification of 631K tables which were extracted from Wikipedia, news sites, etc. All amount of tables were divided into two classes – *main* and *secondary*. The main classification includes *horizontal web table* (Fig. 2, a), *vertical web table* (Fig. 2, b) and *matrix web table* (Fig. 2, c). The secondary classification is represented by following types (generally, tables can belong to several types at the same time):

- *Concise web table* contains merged cells (Fig. 2, d).

Country	Capital	Population
Russian	Moscow	144.1 million
Japan	Tokyo	125.8 million
Germany	Berlin	83.24 million

a

Car	Honda Fit	Kia Rio
Year	2009	2013
Mileage	215 000	114 000
Price	560 000	700 000

b

Studio album	
Released	13 July 1973
Studio	De Lane Lea
Genre	Hard rock

c

Troop strength	1871	1904
Russian	700 000	900 000
Italy	334 000	278 000

d

Events	Monday	Tuesday
May	Trip	Job
June	Learn	Job

e

A a	B b	C c	D d	E e	F f
-----	-----	-----	-----	-----	-----

f

Login	
Password	

g

Figure 1: Examples of tables layout according Crestan classification: vertical listing (*a*); horizontal listing (*b*); attribute/value (*c*); matrix (*d*); calendar (*e*); enumeration (*f*); form (*g*)

- *Split web table* has sequential ordered repeating labels (Fig. 2, *e*).
- A table is classified as *nested web table* if it forms a part of another table (Fig. 2, *f*).
- *Multivalued web table*. Table in which cells have multiple values.
- *Simple multivalued web table* has multivalued data related to one specific type (Fig. 2, *g*).
- *Composed multivalued web table* contains multivalued data related to different types (Fig. 2, *h*).

In the paper [6] the classifying on six types for tables — *layout (non-genuine)*, *genuine* which are divided into *vertical listings*, *horizontal listings*, *matrix*, *others* (all tables that do not belong to the above types) — was suggested. This classifier was tested on a July 2014 version of the Common Crawl.

In 2015, Braunschweig [7] presented several types of tables based on the study of taxonomies in [4, 5], such as *vertical listing* (Fig. 1, *a*), *horizontal listing* (Fig. 1, *b*), *matrix* (Fig. 1, *d*), and special case including *attribute/value* (Fig. 1, *c*) and *nested* (Fig. 2, *f*). In turn, Lehmborg et al. using the WebDataCommons framework [8] extracted and classified tables from the July 2015 version of the Common Crawl based on [7]. As a result, *relational tables* (Fig. 3, *a*) that contain a certain number of entities were extracted. In turn, these entities can be described by attributes, which are represented by rows or columns. Another version of tables in which a certain set of attributes describes one entity only (Fig. 3, *b*) is named *entity tables*. There are also *matrix tables* (Fig. 3, *c*).

The approach of Nishida [9] uses on combines recurrent neural network (RNN) and convolutional neural network (CNN) for table extraction from a subset of the April 2016 Common Crawl. The extracted tables were classified as: *relational tables* (Fig. 4, *a*), *entity tables* (Fig. 4, *b*), *matrix tables* (Fig. 4, *c*) and *other tables*. Tables are represented by semantic triples: subject, property, and object. It is noteworthy that in this taxonomy, *matrix tables* do not contain properties in their structure.

Gasemi-Gol and Szekely presented TabVec [10] — an unsupervised method of embedding tables into a vector space in order to perform classification of tables into categories. Its taxonomy includes *relational* (Fig. 5, *a*), *entity* (Fig. 5, *b*), *matrix* (Fig. 5, *d*) tables. *List* (fig. 5, *c*) and *non-data*.

Year	Title
1869	War and Peace
1878	Anna Karenina
1884	The Decembrists

a

Leo Tolstoy	
Born	9 September 1828
Died	20 November 1910
Language	Russian

b

Oncology	2020	2021
Breast	2.3 M	2.4 M
Stomach	1.1 M	1.1 M

c

Plant	Color	Height
Shrubs		
Azalea	Variable	Shrub
Buddleia	Blue, pink	Shrub

d

Year	Title
2010	Tron
	Trust
	Tangled

Rank	Country	Pop	Rank	Country	Pop
1	China	1.5 B	3	USA	333 M
2	India	1.4 B	4	Indonesia	271 M

e

General information	
Title	Inception
Developed by	Christofer Nolan
Year	2010
Broadcast	
Language	English
Budget	160 M

f

Publisher	Marvel Comics
Title	The Amazing Spider-Man
Main character(s)	Spider-Man, Harry Osborn

g

Albert Einstein	
Born	14.03.1879 German Empire
Doctoral advisor	Alfred Kleiner

h

Figure 2: Examples of tables layout according Lautert classification: horizontal web table (*a*); vertical web table (*b*); matrix web table (*c*); concise web table (*d*); splitted web table (*e*); nested web table (*f*); simple multivalued web table (*g*); composed multivalued web table (*h*)

Index	Weight	Height
21	50	153
19	45	155

a

21	
Weight	50
Height	153

b

Body mass index	45	57
160	Low weight	Normal weight
174	Low weight	Low weight

c

Figure 3: Examples of tables layout according Lehmborg classification: relational tables (*a*); entity tables (*b*); matrix tables (*c*)

The classifier was evaluated on a July 2015 version of the Common Crawl. The same types of tables are also used in the paper [12].

In the [14] Roldan introduced classification of tables types as the following: *horizontal listings* (Fig. 6, *a*), *vertical listings* (Fig. 6, *b*), *form* (Fig. 6, *c*), *matrix* (Fig. 6, *d*). The sources of the tables were Wikipedia (1496 tables) and Dresden Web Table Corpus (1513 tables).

Name	Age	Sex
Anna	25	Fermale
Mike	35	Male

a

Anna	
Age	25
Sex	Fermale

b

Weight

	2020	2021
Anna	57	60
Mike	75	75

c

Figure 4: Examples of tables layout according Nishida classification: relational tables (*a*): contain complete semantic triples (may contain multiple subjects). Anna, Mike – subject, Age, Sex – properties, other – objects; entity tables (*b*): describe the properties of one specific subject only (Anna - subject); matrix tables (*c*): have the same property for each cell object at the junction of a row and a column. Anna, Mike, 2020,2021 – subject, Weight - properties, other – objects

Price	Dish
200	Soup
400	Steak

a

Steak	
Weight	500
Price	400

b

Idaho	
Iowa	
Alabama	

c

Dish

Dish	Soup	Steak
Price	200	400
Weight	400	500

d

Figure 5: Examples of tables layout according Gasemi-Gol classification: relational tables (*a*); entity tables (*b*); list tables (*c*); matrix tables (*d*)

SKU	Item	Price
AU-12	Bread	\$0.90
PZ-18	Butter	\$5.00
XX-99	Water	\$1.00
WI-09	Milk	\$3.95

a

Maker	Apple	Xiaomi
Model	8 Plus	MI6
Screen	LED	LED
RAM	64 Gb	64 Gb
Cores	8	8

b

Name

Name	Pedro
Surname	Lopez
Age	47
Birthplace	Spain

c

	2018	2019
DINER	95B	100B
CHASE	98B	103B
VISA	78B	82B

d

Figure 6: Examples of tables layout according Roldan classification: horizontal listing (*a*); vertical listing (*b*); vertical form (*c*); matrix (*d*)

Table 1
Matching Terminology of Table Type Taxonomies

<i>Crestan</i>	<i>Lautert</i>	<i>Eberius</i>	<i>Braunschweig</i>	<i>Lehmberg</i>	<i>Nishida</i>	<i>Gasemi-Gol</i>	<i>Wang</i>	<i>Roldan</i>
Vertical listing	Horizontal web table	Vertical listing	Vertical listing	(Horizontal) Relational	(Vertical) Relational	Relational	(Vertical) Relational	Horizontal listing
Horizontal listing	Vertical web table	Horizontal listing	Horizontal listing	(Vertical) Relational	(Horizontal) Relational	-	(Horizontal) Relational	Vertical listing
Matrix	Matrix web table	Matrix	Matrix	Matrix	Matrix	Matrix	Matrix	Matrix
Attribute/value	Vertical web table	-	Attribute/value	Entity	Entity	Entity	Entity	Form
Enumeration	-	-	-	-	-	List	List	-

Table 2
Comparison of Table Type Taxonomies

Taxonomy of	Number of types	Levels	Classifier	Classifier availability	Corpora	Corpora availability
Crestan	13	Multi	TabEx (ML-based)	-	(8.2B tables from high quality English pages on the Web – Proprietary crawl)	-
Lautert (main)	7	Multi	WTClassifier (DL-based)	-	(0.6M HTML tables extracted from Wikipedia, e-commerce, news and university sites)	-
Lautert (secondary)	6	Multi	see above	-	see above	-
Eberius	6	Multi	DWTC-Tools ¹ (ML-based)	+	Dresden Web Table Corpus (DWTC) ¹ (125M tables from July 2014 version of the Common Crawl ²)	+
Lehmberg	4	Multi	WDC Extraction Framework ³ (ML-based)	+	Web Data Commons – Web Table Corpora ⁴ (233M tables from July 2015 version of the Common Crawl ²)	+
Nishida	5	Single	TabNet ⁵ (DL-based)	+	(0.19M tables from Top 500 websites from April 2016 version of the Common Crawl ²)	-
Gasemi-Gol	5	Single	TabVec ⁶ (DL-based)	+	(Web pages from unusual domains: human tracking advertisements, fire arms trading, microcap stock market; Random samples from July 2015 Common Crawl ²)	-
Roldan	4	Single	TOMATE ⁷ (cluster analysis)	+	Wikipedia, DWTC ¹	+
Wang	5	Single	TUTA ⁸ (DL-based)	+	(tables extracted from July 2015 version of the Common Crawl ²)	-

3. Results of the Comparison

As can be seen, some of the taxonomies considered above are similar but have different names. It should be noted that there is no any agreement in the literature regarding the types of web-tables. The comparison of web table types of different authors is presented in table 1. Each row of the table contains names for one type of table (if one is presented by the researcher).

The characteristics of the taxonomy kinds and classifiers, such as the number of types, the presence of a hierarchy, the classifier the data set on which the experiments were conducted and their availability are shown in table 2. The column “Number of Types” contains information about the quantity of genuine web-table types proposed by the author. The “Taxonomy levels” column indicates the way for table types definition. Two cases are possible:

¹<https://wwwdb.inf.tu-dresden.de/research-projects/dresden-web-table-corpus/>

²<https://commoncrawl.org/>

³<http://webdatacommons.org/framework/>

⁴<http://webdatacommons.org/webtables/#results-2015>

⁵<https://github.com/dreamquark-ai/tabnet>

⁶<https://github.com/majidghgol/tabvec>

⁷<https://data.mendeley.com/datasets/zcn6h2fvz7/2>

⁸https://github.com/microsoft/TUTA_table_understanding

- Multi-Level – table types are defined hierarchically. Firstly, the set of tables divides into genuine and non-genuine. Next, the genuine tables divide into some types. The obtained types can be divided into other types and so on;
- Single-Level – table types are represented a plain structure. Each table type has no further splitting on other types.

The “Classifier” column shows what classifier and what kind of model are used for types classification. The general, method of classification (ML – “traditional” Machine Learning; DL – Deep Learning; cluster analysis) is given in parentheses. The availability of the classifier is specified in the corresponding column. “Corpora” column contains data about corpora where the tables were extracted from. Information about corpora availability is given in the rightmost column.

4. Summary

In this paper, we explored the various taxonomies of table types that have been proposed in studies of web tables since the groundbreaking work by Cafarella in 2008. Building upon this work, several other research have been conducted, offering different classifications of tables found in HTML documents.

The most comprehensive taxonomy, which includes 12 types, was suggested by Crestan. All other taxonomies we found are based on this one. These taxonomies usually focus on the most common types, such as *listing*, *matrix*, and *attribute/value*. Taking Crestan’s taxonomy as a foundation, Lautert proposed two classes – the main one (containing the basic types of table layouts) and the secondary one (more complex cases of layouts). However, a single table may relate to types in both classes.

Crestan’s primary taxonomy was based on a large corpus of 15 billion HTML tables. The other taxonomies were studied on open corpora with smaller tables amount. Other researchers proposed classifiers based on supervised learning methods.

We have noticed that some of the existing taxonomies are similar but have different terminology. In this study, we have made an attempt to unify the existing terminology by matching different names for common types. Also, the analysis of the taxonomies by comparing their key features was made.

Acknowledgment

The research was supported by the Program of the Fundamental Research of the Siberian Branch of the Russian Academy of Sciences, project no. 121030500071-2

References

- [1] Y. Wang, J. Hu, Detecting tables in html documents, in: International Workshop on Document Analysis Systems, Springer, 2002, pp. 249–260.

- [2] D. Ritze, O. Lehmborg, C. Bizer, Matching html tables to dbpedia, in: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, 2015, pp. 1–6.
- [3] M. Yakout, K. Ganjam, K. Chakrabarti, S. Chaudhuri, Infogather: entity augmentation and attribute discovery by holistic matching with web tables, in: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, 2012, pp. 97–108.
- [4] E. Crestan, P. Pantel, Web-scale table census and classification, in: Proceedings of the fourth ACM international conference on Web search and data mining, 2011, pp. 545–554.
- [5] L. R. Lautert, M. M. Scheidt, C. F. Dorneles, Web table taxonomy and formalization, ACM SIGMOD Record 42 (2013) 28–33.
- [6] J. Eberius, K. Braunschweig, M. Hentsch, M. Thiele, A. Ahmadov, W. Lehner, Building the dresden web table corpus: A classification approach, in: 2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC), IEEE, 2015. URL: <https://doi.org/10.1109/bdc.2015.30>. doi:10.1109/bdc.2015.30.
- [7] K. Braunschweig, Recovering the semantics of tabular web data, Ph.D. thesis, Saechsische Landesbibliothek-Staats-und Universitaetsbibliothek Dresden, 2015.
- [8] O. Lehmborg, D. Ritze, R. Meusel, C. Bizer, A large public corpus of web tables containing time and context metadata, in: Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp. 75–76.
- [9] K. Nishida, K. Sadamitsu, R. Higashinaka, Y. Matsuo, Understanding the semantic structures of tables with a hybrid deep neural network architecture, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.
- [10] M. Ghasemi-Gol, P. Szekely, Tabvec: Table vectors for classification of web tables, arXiv preprint arXiv:1802.06290 (2018).
- [11] J. C. Roldán, P. Jiménez, P. Szekely, R. Corchuelo, Tomate: A heuristic-based approach to extract data from html tables, Information Sciences 577 (2021) 49–68.
- [12] Z. Wang, H. Dong, R. Jia, J. Li, Z. Fu, S. Han, D. Zhang, Tuta: Tree-based transformers for generally structured table pre-training, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 1780–1790.
- [13] M. J. Cafarella, A. Halevy, D. Z. Wang, E. Wu, Y. Zhang, Webtables: exploring the power of tables on the web, Proceedings of the VLDB Endowment 1 (2008) 538–549.
- [14] J. C. Roldán Salvador, Enterprise Data Integration: On Extracting Data from HTML Tables, Ph.D. thesis, 2020.

Literature Review – Ontology Frameworks for Knowledge Management in Software Engineering

Igor Vecstejn¹, Zeljko Stojanov¹, Mila Kavalic¹, Verica Gluvakov¹, Maja Gaborov¹

¹University of Novi Sad, Technical faculty “Mihajlo Pupin”, Djure Djakovica bb, Zrenjanin, Serbia

Abstract

This literature review delves into the domain of Ontology frameworks for Knowledge Management in Software Engineering. Through a systematic analysis of selected primary studies, it becomes evident that ontologies play a crucial role in enhancing knowledge-sharing, collaboration, and decision-making processes within the SE domain. The review highlights the significance of precise domain definition and involvement in the development and utilization of ontologies. A total of 108 papers related to ontologies in KM or SE were identified, with 19 specifically focusing on frameworks, underscoring the importance of frameworks. The review process, guided by established methodologies, ensures the relevance and quality of the selected papers for analysis. This study contributes to a deeper understanding of the utilization of Ontology frameworks in KM in SE, emphasizing the diverse approaches and applications of ontologies in the Software Engineering domain.

Keywords

Knowledge Management, Software Engineering, Ontology, Framework

1. Introduction

Knowledge Management (KM) encompasses both practical business applications and academic exploration. It involves the systematic process of creating, sharing, using, and managing knowledge and information within an organization to improve efficiency and achieve strategic objectives. While the instinct to share knowledge is inherent, the formalization of KM within corporations is relatively recent. Initially rooted in commerce and industry, KM discourse has expanded into interdisciplinary realms, attracting scholars from various fields interested in the creating, sharing, and value of knowledge [1].

In the context of Software Engineering (SE), KM plays a crucial role in capturing, organizing, and leveraging the vast amount of knowledge generated during the software life cycle to enhance productivity and foster innovation [2]. The systematization, reuse, and sharing knowledge becomes even more important in software companies that have multiple teams, departments, or divisions that use different approaches, but at some point should cooperate [3]. Great fluctuation in the workforce in software companies may lead to the loss of knowledge (e.g. when experienced engineers leave the company), which strongly points out the importance of KM in the software industry.

Ontologies, on the other hand, provide a structured way to represent knowledge by defining concepts, relationships, and properties within a specific domain. They help identify inconsistencies and gaps and establish a shared vocabulary within a specific domain. This enables effective information sharing by defining fundamental domain concepts and their interrelationships[4]. Gruber[5] defines an ontology as “explicit formal specifications of the terms in the domain and relations among them.”

By utilizing ontologies in Software Engineering, organizations can facilitate better knowledge-sharing, retrieval, and decision-making processes. Frameworks in the domain of KM in SE offer structure approaches and guidelines for implementing effective KM practices within Software

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: igor.vecstejn@tfzr.rs(A.1); zeljko.stojanov@uns.ac.rs(A.2); mila.kavalic@tfzr.rs(A.3); verica.gluvakov@tfzr.rs(A.4); maja.gaborov@tfzr.rs (A.5);

ORCID: 0009-0009-9940-4411(A.1); 0000-0001-6930-5337(A.2); 0000-0001-7053-2334(A.3); 0000-0003-1083-0203(A.4); 0000-0002-3810-6156 (A.5)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.09

Development processes. These frameworks help organizations streamline knowledge processes, enhance collaboration, and improve project outcomes.

This preliminary literature review on Ontology Frameworks for KM in SE contributes to the initial exploration of existing research in the field. By identifying and analyzing a selection of working papers that focus on Ontologies and KM within SE, this study initiates a systematic understanding of the current landscape.

In the Related Work section, a comprehensive review of some literature reviews on the application of Ontologies for KM in SE will be presented. The Methods Section provides details on the literature review process, including the search strategies employed and the inclusion/exclusion criteria used to select relevant studies. The Results and Discussion section presents the findings from the selected primary studies. Finally, the Conclusion provides a summary of the key insights gained and suggests future research directions in Ontology frameworks for KM in SE.

2. Related Works

This section outlines some recent studies that present literature reviews on using ontologies for knowledge management issues in the software engineering domain. Reviewing scientific papers (not necessarily case studies) that deal with ontologies may significantly contribute to developing the research field and indicate challenges for further research.

Junior et al. [6] presented a systematic literature review (SLR) on tools, models, techniques, and practices utilizing ontology as an aid to Distributed Software Development (DSD). The authors proposed the following research questions that guided their review: (1) Which tools, techniques, models, and practices based on ontologies are used in DSD, and (2) which ontologies were proposed in the field of DSD? Totally 37 primary studies were identified and used for constructing the findings. Identified tools, models, and ontologies were classified based on the phase in the software life cycle proposed in SWEBOK [7]. The following ontologies were identified: Software components, OFFLOSC – Ontology for supporting open source development communities, Open source communities, and ONTODISEN – Ontology to support global software development. Based on the literature review 10 models, 22 tools, 11 practices, and 3 techniques that use ontologies were identified. Regarding the classification of ontologies according to software life cycle phases proposed in SWEBOK, most ontologies relate to software processes in general.

Triandini et al. [8] presented a systematic literature review of the role of ontology in modeling knowledge in software development. The findings provide answers to research questions related to the role of ontology in modeling and the challenges of using ontologies in software engineering. Use of ontologies in software modeling includes modeling data in software development, database, and application integration, using vocabulary for describing knowledge, assistance in software testing, resolving problems with change of software requirements, use in Internet of Things (IoT) systems, quality assessment and improvement of software development, detecting change in software systems, and use of patterns in development. The main identified challenges for using ontologies in software engineering are software development improvement, knowledge abstraction, limited knowledge in malware detection, evaluation of used ontologies, management in software development, ontology in the context of software security, and problems in adjusting to project plans in software development.

Osman et al. [9] present a review of ontology-based knowledge management tools for knowledge sharing in organizations, based on the following comparison criteria: motivation, domain knowledge, source of knowledge, types of knowledge, knowledge extraction, knowledge input process, knowledge retrieval process, knowledge sharing technology, source of ontology component, and ontology methodology. A total of 13 primary studies were identified through filtering studies. Identified tools were used in a variety of domains such as software engineering, education, healthcare, building and construction, decision support systems, economics, etc. The main conclusion of the authors is that tools are specific for narrow domains and that humans are the main source of knowledge. The development and use of ontologies are tightly connected to a precise definition of a domain and scope of knowledge. In contrast, the development process might be manual (from scratch), or by using some tools for automating steps in the development process.

Agbaegbu et al. [10] presented a comprehensive review of existing literature on the application of ontologies in cloud computing. The study aims to identify the evolution of ontology in cloud computing dominant areas of ontology application, less explored areas, challenges faced, and future research directions. The methodology includes the SLR approach, where 400 articles were identified, and 35 relevant articles were selected based on set criteria. The study highlights the impact of cloud ontologies in areas such as cloud security, interoperability, resource description, and service discovery. It also discusses the challenges in applying ontologies to enhance business operations in the cloud and multi-cloud environments. The authors' research aims to provide insights into unresolved challenges and suggest potential future directions for researchers in cloud ontology.

In addition, recent research trends suggest the use of ontologies in domains such as systems engineering [11] and systems of systems [12], which are complex systems that commonly include software systems. Further, Kotis and Vouros [13] suggested that continuously exploring and evaluating ontologies is necessary since ontologies may change in the evolving context, which indicates the need for active involvement of all relevant stakeholders during the ontology life cycle.

3. Methods

A more comprehensive literature review related to frameworks, methods, and models in Knowledge Management within Software Engineering has been performed. The initial objective was to investigate the activities and focus areas of individuals in the field. Our research concluded that frameworks, models, and methods are commonly employed to address specific problems. The objective of this paper is to focus solely on the review of frameworks in Knowledge Management in Software Engineering based on Ontologies.

Certain procedures for the literature review used in this study are adapted from [14]. While our study is not a comprehensive Systematic Literature Review, specific methodologies presented in Chapter 22 of the book [14] were employed to guide our Literature Review process, which is shown in Figure 1.

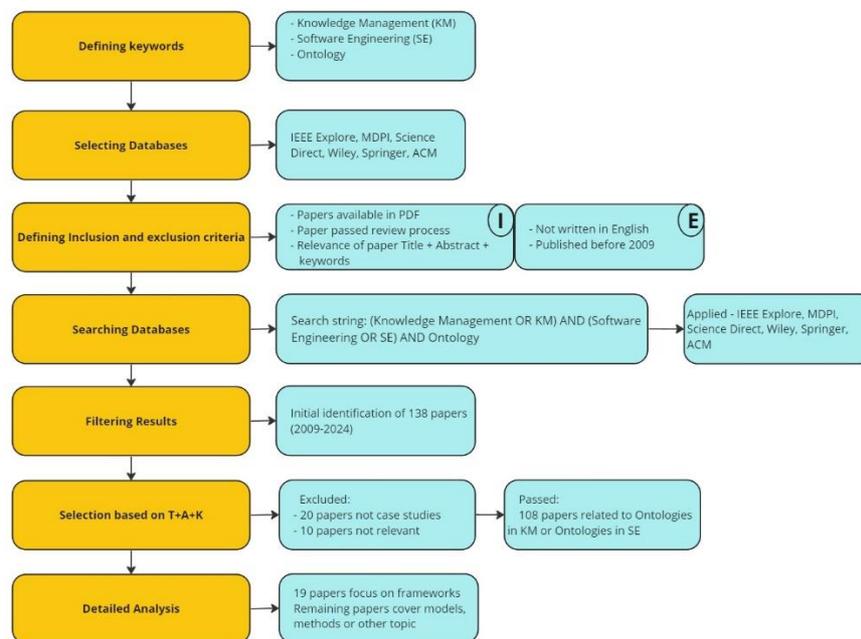


Figure 1: Literature Review process

As seen in Figure 1, yellow blocks show Phases of the literature review, while cyan blocks show the outputs of each phase.

For this preliminary Literature Review, three primary keywords were proposed:

- Keyword 1 (KW1) – “Knowledge Management” OR “KM”
- Keyword 2 (KW2) – “Software Engineering” OR “SE”
- Keyword 3 (KW3) – “Ontology”

These three primary keywords represent essential concepts guiding the search for relevant literature. To enhance the precision of the literature search, logical operators were employed to combine search terms. The Boolean operators AND and OR were utilized to connect key concepts within the search of a string. Combining these three keywords, the following search string was formed:

KW1 AND KW2 AND KW3

This literature search string was applied across databases such as IEEE Explore, MDPI, ScienceDirect, and Wiley. An additional search to find out some papers not published in the stated databases was performed by using the Google Scholar Web Search Engine.

The selection of papers for this preliminary Literature Review adhered to specific inclusion criteria. Proposed inclusion criteria are: (I1) Only papers that are available in PDF format were considered to ensure accessibility; (I2) Papers that passed the review process (peer-reviewed journals and conference proceedings); (I3) The relevance of the paper’s title, abstract, and keywords (T+A+KW) ensuring alignment with the research focus.

To ensure the relevance and quality of the literature selected for this preliminary review certain exclusion criteria were applied. Used exclusion criteria are: (E1) Papers that are not written in English; (E2) Papers that were published before 2009.

Based on the specified search string and the inclusion and exclusion criteria, the search was conducted across selected databases. A total of 138 papers published between 2009 and 2024 were identified. Upon briefly reading the text of the collected research papers, it was concluded that 20 papers were not case studies, and 10 did not correspond to the topic. The remaining 108 papers, which are related to the thematic fields of Ontologies in Knowledge Management or Ontologies in Software Engineering were included for more detailed analysis.

A detailed analysis of these 108 papers revealed that 19 specifically focus on frameworks, while the remaining papers address models, methods, or various other topics. These 19 research papers were selected as primary studies (PS) and are presented in Table 1.

Table 1
Selected primary studies

Primary Studies	Reference	Year
PS1	Ribino, P., Oliveri, A., Re, G. L., & Gaglio, S. (2009, June). A knowledge management system based on ontologies. In 2009 International Conference on New Trends in Information and Service Science (pp. 1025-1033). IEEE.	2009
PS2	Tuan, D. T., & Tuan, D. C. (2011, October). Enhance Java Software Development with Knowledge Acquisition and Management Tools. In 2011 Third International Conference on Knowledge and Systems Engineering (pp. 70-77). IEEE.	2011
PS3	Zheng, Y. L., He, Q. Y., Ping, Q. I. A. N., & Ze, L. I. (2012). Construction of the ontology-based agricultural knowledge management system. <i>Journal of Integrative Agriculture</i> , 11(5), 700-709.	2012
PS4	Moser, T., & Biffl, S. (2011). Semantic integration of software and systems engineering environments. <i>IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)</i> , 42(1), 38-50.	2011
PS5	Gobin, B. A. (2012, November). A domain ontology for capturing knowledge for social integration. In 2012 Tenth International Conference on ICT and Knowledge Engineering (pp. 194-200). IEEE.	2012
PS6	Jin, F., Zhang, X., & Liu, W. (2014, October). Enterprise Knowledge Management Platform Based on Fuzzy Ontology. In 2014 International Conference on Management of e-Commerce and e-Government (pp. 246-251). IEEE.	2014

PS7	Ramona-Cristina, P., Vasileanu, A., & Goga, N. (2016, October). Ontology-based multi-system for SME knowledge workers. In <i>2016 IEEE International Symposium on Systems Engineering (ISSE)</i> (pp. 1-5). IEEE.	2016
PS8	Knauss, E., Liebel, G., Schneider, K., Horkoff, J., & Kasauli, R. (2017, September). Quality requirements in agile as a knowledge management problem: More than just-in-time. In <i>2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)</i> (pp. 427-430). IEEE.	2017
PS9	Yuan, E. (2017, May). Towards ontology-based software architecture representations. In <i>2017 IEEE/ACM 1st International Workshop on Establishing the Community-Wide Infrastructure for Architecture-Based Software Engineering (ECASE)</i> (pp. 21-27). IEEE.	2017
PS10	Alsanaad, A. A., Chikh, A., & Mirza, A. (2019). Multilevel ontology framework for improving requirements change management in global software development. <i>IEEE Access</i> , 7, 71804-71812.	2019
PS11	Abioye, T. E., Arogundade, O. T., Misra, S., Akinwale, A. T., & Adeniran, O. J. (2020). Toward ontology-based risk management framework for software projects: an empirical study. <i>Journal of Software: Evolution and Process</i> , 32(12), e2269.	2020
PS12	Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., ... & Schuecker, J. (2021). Informed machine learning—a taxonomy and survey of integrating prior knowledge into learning systems. <i>IEEE Transactions on Knowledge and Data Engineering</i> , 35(1), 614-633.	2021
PS13	Bernasconi, E., Di Pierro, D., Redavid, D., & Ferilli, S. (2023). SKATEBOARD: Semantic Knowledge Advanced Tool for Extraction, Browsing, Organisation, Annotation, Retrieval, and Discovery. <i>Applied Sciences</i> , 13(21), 11782.	2023
PS14	Uwasomba, C. F., Lee, Y., Yusoff, Z., & Chin, T. M. (2022). Ontology-Based Methodology for Knowledge Acquisition from Groupware. <i>Applied Sciences</i> , 12(3), 1448.	2022
PS15	Elhabbash, A., Elkhathib, Y., Nundloll, V., Marco, V. S., & Blair, G. S. (2024). Principled and automated system of systems composition using an ontological architecture. <i>Future Generation Computer Systems</i> .	2024
PS16	Hasan, N., Rahman, M. H., Wessman, A., Smith, T., & Shafae, M. (2023). Process Defects Knowledge Modeling in Laser Powder Bed Fusion Additive Manufacturing: An Ontological Framework. <i>Manufacturing Letters</i> , 35, 822-833.	2023
PS17	Zhang, L., Olsen, A., & Lobov, A. (2022). An ontology-based KBE application for supply chain sustainability assessment. <i>Resources, Environment and Sustainability</i> , 10, 100086.	2022
PS18	Nazir, S., Motla, Y. H., Abbas, T., Khatoon, A., Jabeen, J., Iqra, M., & Bakhat, K. (2014, November). A process improvement in requirement verification and validation using ontology. In <i>Asia-Pacific World Congress on Computer Science and Engineering</i> (pp. 1-8). IEEE.	2014
PS19	Ghaisas, S., & Ajmeri, N. (2013). Knowledge-assisted ontology-based requirements evolution. In <i>Managing requirements knowledge</i> (pp. 143-167). Berlin, Heidelberg: Springer Berlin Heidelberg.	2013

4. Results and Discussions

From the primary studies, through detailed reading and analysis of research papers, 11 types of frameworks were identified and presented in Table 2.

Table 2
Types of frameworks

Types of Frameworks	Primary Studies
Knowledge Management FW	PS1, PS3, PS4, PS6, PS8, PS10, PS13, PS16
Software Development FW	PS2
Ontology Development FW	PS5
Information Utilization FW	PS7
Software Architecture Modeling FW	PS9
Risk Management FW	PS11
Machine Learning Taxonomy FW	PS12
Knowledge Harvesting FW	PS14
System of Systems FW	PS15

Knowledge Representation FW	PS17
Ontology-based Requirements Engineering FW	PS18, PS19

In the realm of KM, various frameworks have been developed and utilized to enhance the organization and utilization of knowledge. The primary studies that discuss or utilize knowledge management framework include [PS1][PS3][PS4][PS6][PS8][PS10][PS13][PS16]. Studies [PS18][PS19] explore ontology-based requirements engineering frameworks, which utilize ontological approaches to enhance the requirement engineering process. In addition to these, there are frameworks specifically focused on Software Development, Ontology development, Information Utilization, Software Architecture Modeling, Risk Management, Machine Learning Taxonomy, Knowledge Harvesting, System of Systems, and Knowledge Representation.

From primary studies, various domains that focus on the development or utilization of frameworks were identified. These domains, along with their corresponding primary studies are presented in Table 3.

Table 3
Categorized domains based on primary studies

Domains	Primary Studies
Additive Manufacturing	PS16
Knowledge Management	PS1, PS3, PS6, PS8
Distributed Systems	PS15
Engineering	PS4
Knowledge Acquisition	PS14
Semantic Technologies	PS13
Knowledge	PS5
Machine Learning	PS12
Product Lifecycle Management	PS17
Requirements Engineering	PS18, PS19
Software Engineering	PS2, PS9, PS10, PS11
Software Implementation for SMEs	PS7

Knowledge Management frameworks are a significant area of focus, with primary studies [PS1][PS3][PS6][PS8]. Software Engineering is another major domain with [PS2][PS9][PS10][PS11]. In the field of Requirements Engineering, there are studies [PS18] and [PS19]. There are also other domains such as Additive Manufacturing, Distributed Systems, Engineering in general, Knowledge Acquisition, Semantic Technologies, Knowledge, Machine Learning, Product Lifecycle Management, and Software Implementation for SMEs.

4.1. Constraints and validity

Literature reviews in software engineering need to adjust to validity criteria to properly address identified threats and increase the reliability and usability of reported findings [15]. The authors are aware of potential threats to validity, such as the construction of a search string, selection of databases to search, assessment of the quality of selected papers, and construction of the findings. Discussion of two important aspects of validity, named internal and external validity contributes to the soundness of the research process and reliability of the findings.

Presented details on planning and conducting a literature review are essential for addressing internal validity, which is done in the third section of the paper. Based on the description of the review process,

including the search for papers and their analysis, enables traceability from findings to the selected primary studies. This increases the reliability of the presented findings. External validity relates to the generalizability of the findings. Since the findings are specific to the stated field, there is no intention to generalize findings to other fields in software engineering. However, the authors suggest that the presented literature review process can be generalized and applied in other fields.

5. Conclusion

This preliminary Literature Review has provided valuable insights into the utilization of Ontology frameworks for KM in SE. Through the analysis of selected primary studies, it is evident that ontologies play a crucial role in structuring and organizing information in various domains, including Software Engineering, Education, Healthcare, and Decision Support Systems. The review highlighted the importance of precise domain definition and human involvement in the development and use of Ontologies.

Furthermore, the study identified a total of 108 papers related to ontologies in KM or SE, with 19 specifically focusing on frameworks. This underscores the significance of frameworks in addressing specific problems within the field. The review process, guided by methodologies adapted from existing literature, ensured the relevance and quality of the selected papers for analysis.

This paper contributes to the understanding of the role of Ontological frameworks in KM in SE through the analysis of relevant literature. The identification of key words and the classification of framework types provide insight into the diversity of approaches and applications of Ontologies in the domain of SE. Through the analysis of selected studies, the paper emphasized the importance of ontological frameworks in organizing information, facilitating collaboration, and improving the decision-making process in SE.

Looking ahead, the insights and findings from this review lay the groundwork for future investigations into Ontology frameworks for KM in SE. By addressing existing research gaps and exploring emerging trends, researchers can further enhance the effectiveness of Ontologies in facilitating KM processes within the SE domain. Future research could focus on the development, implementation, and evaluation of Ontology frameworks tailored to specific domains within SE, or other fields of human work and life. These frameworks could address evolving challenges in KM and offer innovative solutions for information organization.

6. References

- [1] McInerney, C. (2002). Knowledge management and the dynamic nature of knowledge. *Journal of the American society for Information Science and Technology*, 53(12), 1009-1018
- [2] Schneider, K., *Experience and Knowledge Management in Software Engineering*. Springer-Verlag, Berlin, Heidelberg, 2009. doi:10.1007/978-3-540-95880-2.
- [3] Ivarsson, M., Gorschek, T., 2012. Tool support for disseminating and improving development practices. *Software Quality Journal*, 20(1), pp.173–199. doi:10.1007/s11219-011-9139-6.
- [4] R. Gayathri and V. Uma, "Ontology-based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey," *ICT Express*, vol. 3, no. 2, pp. 69_74, 2018.
- [5] Gruber, T.R., 1993. A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition*, 52(6), pp.1111-1133.
- [6] Junior, A.N.B., de Azevedo, R.R., da Silva, F.Q., Rocha, R.G. and Costa, C., 2012. Ontologies Supporting the Distributed Software Development: A Systematic Literature Review. In *2012 IEEE Seventh International Conference on Global Software Engineering Workshops* (pp. 55-59). IEEE.
- [7] Bourque, P. & Fairley, R. E., ed. (2014), *SWEBOK: Guide to the Software Engineering Body of Knowledge*, IEEE Computer Society, Los Alamitos, CA
- [8] Triandini, E., Kristyanto, M.A., Rishika, R.V., &Rawung, F. (2021). A Systematic Literature Review of the Role of Ontology in Modeling Knowledge in Software Development Processes. *IPTEK The Journal The Journal for Technology and Science*, 32(3), 159-175

- [9] Osman, M.A., Noah, S.A.M. and Saad, S., 2022. Ontology-based knowledge management tools for knowledge sharing in organization—a review. *IEEE Access*, 10, pp.43267-43283.
- [10] Agbaegbu, J., Arogundade, O.T., Misra, S., & Damaševičius, R. (2021). Ontologies in cloud computing- review and future directions. *Future Internet*, 13(12),302
- [11] L. Yang, K. Cormican, M. Yu, Ontology-based systems engineering: A state-of-the-art review, *Computers in Industry* 111 (2019) 148-171. doi:10.1016/j.compind.2019.05.003.
- [12] G. Abdalla, C. D. N. Damasceno, M. Guessi, F. Oquendo, E. Y. Nakagawa, A Systematic Literature Review on Knowledge Representation Approaches for Systems-of-Systems, in: *Proceedings of the 2015 IX Brazilian Symposium on Components, Architectures and Reuse Software (SBCARS '15)*, 2015, pp. 70–79. doi:10.1109/SBCARS.2015.18.
- [13] K. I. Kotis, G. A. Vouros, D. Spiliotopoulos, Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations, *The Knowledge Engineering Review* 35 (2020) e4. doi:10.1017/S0269888920000065.
- [14] Kitchenham, B.A., Budgen, D., & Brereton, P. (2015). *Evidence-based software engineering and systematic reviews (Vol.4)*. CRC press
- [15] X. Zhou, Y. Jin, H. Zhang, S. Li and X. Huang, A Map of Threats to Validity of Systematic Literature Reviews in Software Engineering, in: *Proceedings of the 23rd Asia-Pacific Software Engineering Conference (APSEC)*, 2016, pp. 153-160, Hamilton, New Zealand. doi:10.1109/APSEC.2016.031.

Методика построения аналитической модели потребления природного газа

Ольга Башарина^{1,2}, Алексей Еделев^{1,3}, Татьяна Фер²

¹ Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

² Уральский государственный экономический университет, ул. 8 марта, д. 62, 620144, Екатеринбург, Россия

³ Институт систем энергетики им. Л.А. Мелентьева СО РАН, ул. Лермонтова, д. 130, 664033, Иркутск, Россия

Аннотация

Представлена методика построения математической модели на основе анализа данных временных рядов потребления природного газа на территории Евросоюза. Методика включает три основных этапа: предварительный анализ данных; построение тренд-сезонной модели; оценка качества и точности модели. Использование методики продемонстрировано на примере и показывает хорошее качество и точность полученной аналитической модели.

Ключевые слова

Временной ряд, тренд-сезонная модель, гармонический анализ

1. Введение

Необходимость в разработке новых моделей систем газо-, нефте- и нефтепродуктоснабжения для исследований живучести основана, в частности, на однопериодности разработанных в ИСЭМ СО РАН моделей этих больших трубопроводных систем [1, 2, 3]. Следствиями этого недостатка являются отсутствие динамики и упрощенное моделирование экстремальных условий функционирования систем энергетики [3].

Новые модели систем газо-, нефте- и нефтепродуктоснабжения позволят более реалистично и точно моделировать функционирование этих систем энергетики в экстремальных условиях. При том серьезной проблемой является отсутствие в открытом доступе исходных данных, необходимых для построения многопериодных моделей больших трубопроводных систем, в частности для получения посуточных графиков потребления энергоресурсов. В [1] показан подход к решению задачи построения схемы газоснабжения Европейского союза на основе открытых данных, являющихся результатом проекта SciGRID_gas [4] Ольденбургского института сетевых энергетических систем.

В рамках проекта SciGRID_gas были получены и опубликованы массивы данных SciGRID_gas CONS [5] ежесуточного потребления природного газа за 10 лет для всех регионов стран Евросоюза на уровнях NUTS-1 (федеральный уровень), NUTS-2 (административные округа) и NUTS-3 (районы и города без районов).

Цель нашего исследования – разработать методику построения модели потребления газа для территориальных субъектов Евросоюза уровня NUTS-3.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: basharinaolga@mail.com (A. 1); flower@isem.irk.ru (A. 2); tatiana.fer@yandex.ru (A. 3)

ORCID: 0000-0002-7151-782X (A. 1); 0000-0003-3034-4907 (A. 2)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.10

2. Материалы и методология

Исходными данными послужили показатели ежедневного потребления природного газа в 1364 городах и районах стран Евросоюза за период 2010–2019 гг. с детализацией по промышленному, жилому и коммерческому секторам. В рамках нашего исследования интерес представляет суммарное потребление газа, таким образом исходные данные по каждому территориальному субъекту представлены в виде одномерного временного ряда, содержащего 3652 уровня.

Методика построения аналитической модели включает три основных этапа:

1. Предварительный анализ данных.
2. Построение тренд-сезонной модели.
3. Оценка адекватности и точности модели.

Предварительный анализ ряда проводится на основе графического представления анализируемых данных ежедневного потребления природного газа, а также с использованием статистических критериев. На данном этапе выполняется начальное описание ряда, предварительная обработка данных, включая устранение аномальных наблюдений, определение стационарности ряда.

Визуальный анализ графика временного ряда позволяет выдвинуть предположение, о том что ряд содержит трендовую $T(t)$, сезонную $S(t)$ и случайную $E(t)$ составляющие. Поскольку амплитуды колебаний уровней ряда примерно одинаковые, для моделирования ряда выбрана аддитивная модель:

$$Y(t) = T(t) + S(t) + E(t). \quad (1)$$

Проверка наличия тенденции и определение периодичности сезонных колебаний во временном ряде выполняется на основе анализа автокорреляционной функции уровней ряда и коррелограммы [6]. Такой анализ позволяет найти величину лага, при котором автокорреляция уровней ряда наиболее высокая.

Если наиболее высоким оказывается коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию. Если наиболее высоким оказывается коэффициент автокорреляции порядка p , то ряд содержит колебания с периодичностью в p моментов времени.

Для исследуемых временных рядов структура оказалась следующей: определено наличие линейного тренда и сезонности с периодом 365 дней.

Следующий этап представленной методики заключается в построении тренд-сезонной модели. При анализе графика временного ряда был сделан вывод о линейном характере тренда. Методом гармонического (спектрального) анализа, позволяющего представить периодические сезонные колебания в виде линейной комбинации синусов и косинусов, была определена гармоника под номером 1 с наиболее значимым спектром [6-7].

Таким образом, модель ежедневного потребления природного газа представляется суммой линейного тренда, сезонной компоненты с периодом 365, описываемой тригонометрическими функциями, и случайной составляющей:

$$Y(t) = a_0 + a_1 t + a_2 \cos \frac{2\pi t}{365} + a_3 \sin \frac{2\pi t}{365} + E(t), \quad (2)$$

где t – уровень ряда; a_0, a_1, a_2, a_3 – параметры уравнения регрессии, рассчитанные методом наименьших квадратов; $E(t)$ – случайная составляющая.

На заключительном этапе моделирования проводится оценка адекватности полученной модели по следующим характеристикам.

Коэффициент детерминации R^2 . Измеряет долю дисперсии, объясненную моделью, в общей дисперсии уровней временного ряда. Чем ближе коэффициент детерминации к единице, тем более хорошо полученная аналитическая модель соответствует исходному временному ряду.

F-критерий Фишера. Используется для оценки значимости модели «в целом». Расчетное значение F-критерия сравнивают с табличным при выбранном уровне значимости. Если выполняется неравенство $F_{\text{расч}} > F_{\text{табл}}$, то модель признается значимой, а количество наблюдений достаточным [6-8].

Точность модели оценивают с помощью ошибок аппроксимации. Средняя абсолютная ошибка (MAE, Mean Absolute Error) отражает среднюю абсолютную разницу между модельными (расчетными) и фактическими значениями временного ряда. MAE устойчива к выбросам и не усиливает большие ошибки. Средняя квадратическая ошибка (MSE, Mean Squared Error), напротив, выявляет грубые ошибки за счет того, что величина отклонения фактических уровней от расчетных возводится во вторую степень. На практике из величины MSE извлекают квадратный корень и получают RMSE (Root Mean Squared Error), называемую стандартной ошибкой. Средняя абсолютная процентная ошибка (MAPE, Mean Absolute Percentage Error) показывает среднее абсолютное отклонение расчетных значений от фактических в процентах и имеет простую интерпретацию.

3. Пример реализации методики

В качестве примера реализации представленной методики приведем процедуру построения модели потребления газа для территориального субъекта Греции CZ041.

На этапе предварительного анализа проведена проверка исходных данных, визуальный анализ графика временного ряда. Анализ автокорреляционной функции и коррелограммы временного ряда (рисунок 1) показал наличие в его составе тенденции и сезонной компоненты с периодом 365 дней.

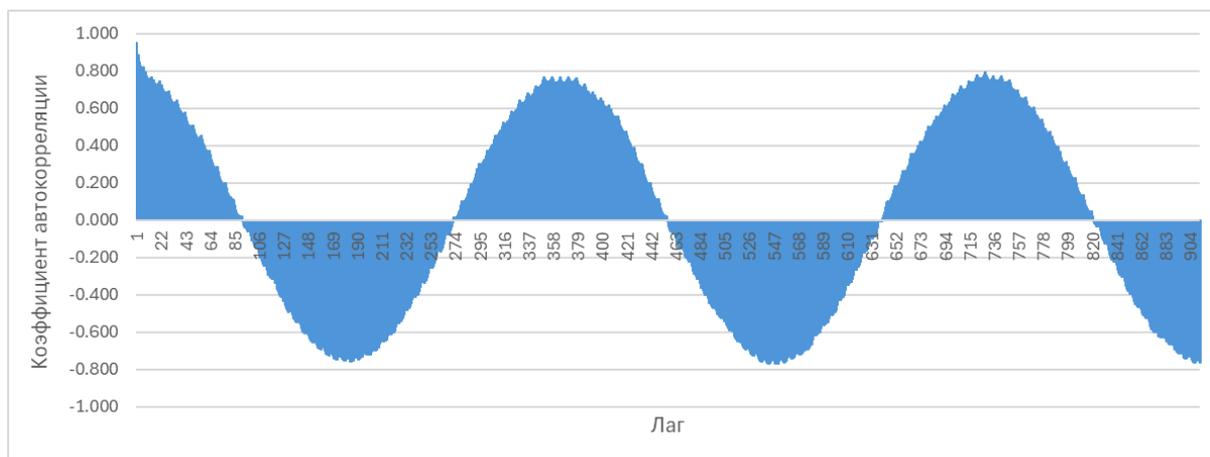


Рисунок 1: Коррелограмма временного ряда

В результате построения тренд-сезонной модели в соответствии с (1) получена оценка ежесуточного потребления газа:

$$\hat{Y}(t) = 0,487 - 0,00001t + 0,176 \cos \frac{2\pi t}{365} + 0,056 \sin \frac{2\pi t}{365}, \quad (2)$$

где t – номер дня.

На рисунке 2 представлен график сравнения фактических и расчетных значений объема потребления газа.



Рисунок 2: Ежесуточный объем потребления газа (фактическое и расчетное значения)

График остаточной компоненты построенной модели представлен на рисунке 3. Судя по графику, остаточная компонента носит случайный характер

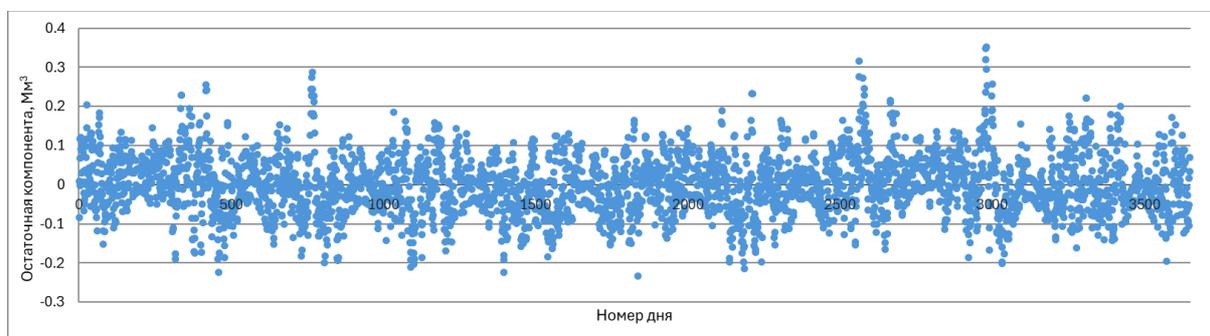


Рисунок 3: Остаточная компонента временного ряда

Коэффициент детерминации $R^2 = 75.8\%$. Построенная тренд-сезонная модель объясняет 75.8% вариации потребления природного газа, а остальные 24.2% приходятся на влияние внешних факторов.

При проверке значимости модели в целом получено расчетное значение F-статистики $F_{расч} = 3813$, значительно превышающее табличное значение $F_{табл} = 2.6$ для уровня значимости 0.05. Модель признается значимой в целом с вероятностью 95% и пригодной для прогнозирования.

При оценке точности модели рассчитаны значения ошибок: $MAE = 0,057 \text{ Мм}^3$; $RMSE = 0.074 \text{ Мм}^3$; $MAPE = 12.4\%$. Точность модели можно считать удовлетворительной.

4. Заключение

Представленная в статье методика позволяет построить тренд-сезонную модель потребления природного газа. Результаты оценки адекватности и точности модели дают возможность использовать ее для анализа и прогнозирования потребления газа для территориальных субъектов Евросоюза уровня NUTS-3. Модель направлена на реализацию в распределенной вычислительной среде для ускорения расчетов по множеству потребителей. Описанная методика частично реализована в виде сервиса.

Дальнейшее направление исследования связано с проведением анализа зависимости потребления газа от природно-климатических показателей.

5. Благодарности

Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды», рег. № 124052100088-3).

6. Список литературы

- [1] А. Еделев, Г. Данилов, О. Башарина, Разработка модели системы газоснабжения на основе открытых данных для исследования живучести, Материалы VI Международного семинара по информационным, вычислительным и управляющим системам для распределенных сред (ICCS-DE 2024, 1-5 июля 2024 г., г. Иркутск, Россия), Иркутск, Изд-во ИДСТУ СО РАН, 2024, С. 103-110.
- [2] А.В. Храмов, С.М. Еникеева Программное и информационное обеспечение решения задач живучести Единой системы газоснабжения СССР, Методы и модели исследования живучести систем энергетики, Новосибирск: Наука, Сиб. Отд, 1990, С. 86-91.
- [3] С.М. Клименко, С.М. Сендеров, В.А. Янченко, Исследование проблем повышения устойчивости и экологической безопасности магистральных нефте- и нефтепродуктопроводов, Новые информационные технологии управления развитием и функционированием трубопроводных систем энергетики, Иркутск: СЭИ СО РАН, 1993, С. 119-127.
- [4] A. Pluta, W. Medjroubi, J.C. Diettrich, J. Dasenbrock, H.P. Tetens, J.E. Sandoval, O. Lünsdorf, SciGRID_gas-Data Model of the European Gas Transport Network, In 2022 Open Source Modelling and Simulation of Energy Systems (OSMSES), 2022, April, pp. 1-7.
- [5] Jan Diettrich, Adam Pluta, Wided Medjroubi. SciGRID_gas IGGIELGN, Version 1.1.2. Zenodo, May 2021. doi: 10.5281/zenodo.4767098.
- [6] В. А. Алексеева, Анализ временных рядов : уч. пособие, Ульяновск : УлГТУ, 2020, 147 с.
- [7] Н. А. Дегтярева, Модели анализа и прогнозирования на основе временных рядов: монография, Челябинск: Изд-во ЗАО «Библиотека А.Миллера», 2018, 160 с.
- [8] Е. В. Радковская, Е. М. Кочкина, М. В. Дроботун, Т. В. Фер, Н. П. Попова, И. В. Иванов, Эконометрика, США, Open Science Publishing, 2019, 177 с., ISBN 978-0-244-47834-6.

Методика выбора компенсирующих мероприятий в исследованиях живучести топливно-энергетического комплекса страны

Наталья Береснева¹, Алексей Еделев^{1,2}

¹ Институт систем энергетики им. Л.А. Мелентьева СО РАН, ул. Лермонтова, д. 130, 664033, Иркутск, Россия

² Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

Аннотация

В статье рассмотрена облегченная методика выбора инвариантных мероприятий по повышению живучести (компенсирующих мероприятий) ТЭК страны. Ключевая особенность данной методики – автоматическое определение инвариантных мероприятий в рамках одноуровневой однокритериальной оптимизации. Ядро данной методики – многопериодная модель функционирования ТЭК с настраиваемой логикой комбинирования нештатных ситуаций (НС) и компенсирующих мероприятий (КМ). Оптимизация выбора мероприятий в ней реализована благодаря:

- двухуровневому временному моделированию аномальных условий функционирования ТЭК по интервалам времени, включающим технологически взаимосвязанные временные периоды;
- настройки логики формирования наборов КМ за счет бинарной активизации мероприятий в наборах, за счет задания логики совместной реализации мероприятий (правила взаимоисключения – логическое «или» и взаимодополнения – логическое «и»);
- задания в целевой функции критерия минимальных суммарных ущербов по всем временным интервалам при минимальных издержках на введение КМ.

Инструментальная реализация методики ориентирована на шаблонное задание КМ по схеме «Тип мероприятий» - «Тип объектов, подверженных данным мероприятиям» - «Критерий отбора объектов в наборы инвариантных мероприятий». Предлагаемая методика применима в условиях ограниченного множества наиболее представительных НС, в условиях достаточности однокритериальной оценки мероприятий. Апробация методики предлагается на примере оценки влияния реконструкции сети Единой систем газоснабжения РФ на конечных потребителей ТЭК страны.

Ключевые слова

Живучесть топливно-энергетического комплекса, нештатные ситуации, компенсирующие мероприятия, модель функционирования ТЭК.

1. Введение

Ключевым фактором обеспечения стабильного функционирования ТЭК страны является обеспечение его живучести, под которой понимается свойство входящих в него систем энергетики противостоять возмущениям (внутренним и внешним), не допуская нарушения энергоснабжения на обширной территории на длительное время, либо сводя к минимуму

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL beresneva@isem.irk.ru (A. 1); flower@isem.irk.ru (A. 2)

ORCID: 0000-0002-0003-947X (A. 1); 0000-0003-2219-9754 (A. 2)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.11

катастрофические последствия таких нарушений [1]. Важная составляющая живучести - способность восстановления ТЭК своей производительности до исходного (до момента возникновения НС) или приближенном к нему объеме.

На практике оценка живучести ТЭК сводится к отслеживанию и сравнению изменений между его состояниями на протяжении времени его выживания. Данная оценка определяется в рамках модельных вычислительных экспериментов с различных точек зрения (функциональной, топологической, экономической), требует многокритериального анализа рассчитанных сводных метрик [2]. Такой вариант исследований универсален в предметном (может быть применим к любой оптимизационной модели ТЭК) и вычислительном (предусматривает высокопроизводительные вычисления большого количества сценариев) отношении. Но он не целесообразен в случаях актуальности быстрых решений по выбору КМ. Для этих случаев предлагается облегченная методика определения наборов инвариантных наборов КМ для аномальных ситуаций функционирования ТЭК с серьезными системными последствиями. Данная методика позволяет на базе оптимизационной многопериодной модели ТЭК автоматически получать наборы инвариантных мероприятий, эффективных для рассмотренных НС. Основные положения этой методики, результаты ее апробации приведены в следующих разделах публикации.

2. Методика однокритериального выбора набора инвариантных компенсирующих

Предлагаемая облегченная методика выбора мероприятий по повышению живучести отвечает следующим ключевым моментам:

1. Облегченный характер методики, ее базирование на многопериодной модели функционирования ТЭК. Особенности модели:

- двухуровневая временная схема задания сценариев исследования - рассматриваются изолированные временные интервалы, включающие временные схемы функционирования ТЭК;
- оптимизация по критерию максимуму обеспечения потребностей при минимальных затратах на ввод КМ по всем временным интервалам;
- применение для оценки КМ трех типов переменных:
- дискретных переменных для идентификация оптимизатором статуса инвариантности мероприятий по всем временным интервалам;
- непрерывных переменных для оценки эксплуатационных затрат мероприятий, их интенсивности по отдельным временным интервалам.

Благодаря этому набор инвариантных КМ определяется в рамках одного вычислительного эксперимента по всем возможным НС.

2. Включение в сценарии исследований набора наиболее критичных НС (каждая НС моделируется на отдельном временном интервале). По КМ – задание избыточного множеств мероприятий определенного типа. Каждая КМ может быть задана графиком ее ввода в эксплуатацию по временным диапазонам для всех временных интервалов. Таким образом по всем временным интервалам создается единый избыточный набор мероприятий. За счет этого достигается фактор инвариантности оптимальных КМ.

3. Учет территориального аспекта в выборе инвариантных мероприятий за счет гибкости задания целевой функции модели (возможности учета штрафов за недопоставку энергоресурсов по различным территориям. Возможность ранжирования мероприятий внутри набора КМ путем анализа удельной эффективности мероприятий. Как следствие, возможность апостериорного анализа множества наборов инвариантных мероприятий при условии задания единого множества НС.

4. Возможность задания логики ввода мероприятий с помощью правил взаимного исключения мероприятий (логическое «или», логическое «и»), позволяющая учесть особенности реализации комплексной реализации мероприятий. Мероприятия, включенные в данные правила, автоматически исключаются по всем временным интервалам путем изменения флага бинарной переменной мероприятий.

Концепция методики однокритериального выбора набора инвариантных КМ приведена на рис. 1. Этапность решения задачи выбора КМ в методике практически не отличается от стандартной схемы проведения вычислительного эксперимента, в рамках отдельного исследования включает следующую последовательность действий:

1. Поэтапное задание множества НС. Для каждой НС автоматически создается новый временной интервал.
2. Выбор типа КМ (например, новые участки магистральных газопроводов), критерия отбора объектов (например, длина вводимого участка магистрального газопровода) в наборы инвариантных мероприятий.
3. Оптимизация потокораспределения многопериодной модели исследования;
4. Вывод выбранного оптимизатором набора инвариантных КМ.
5. Апостериорный анализ полученных результатов.

Математическая формализация многопериодной модели функционирования ТЭК приведена ниже.

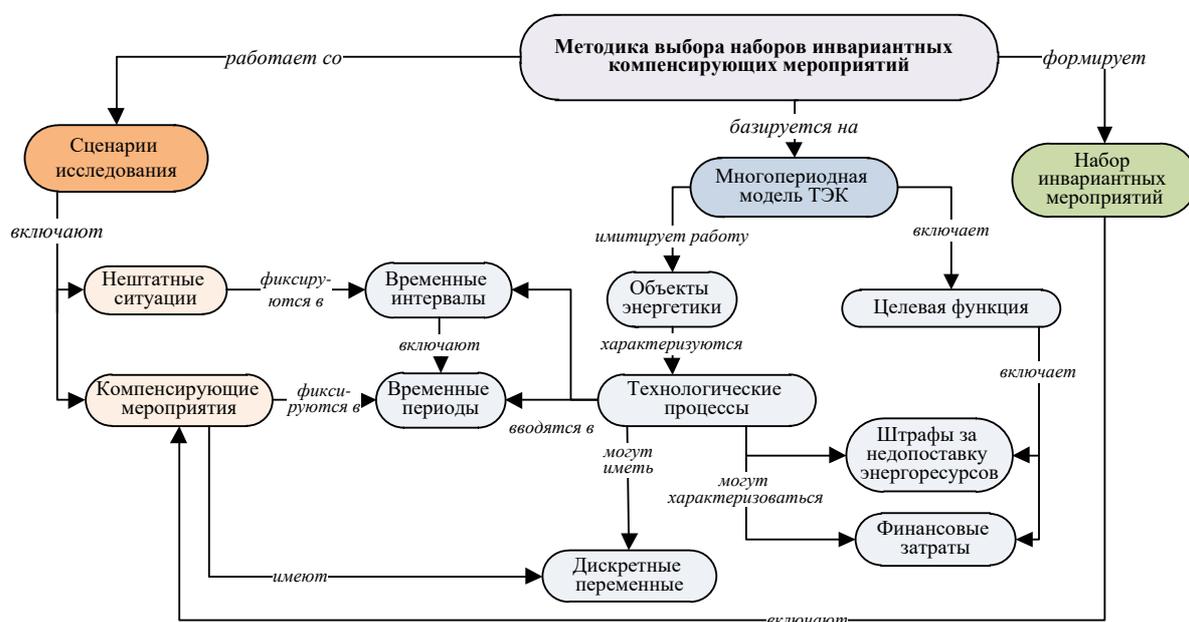


Рисунок 1: Концепция методики однокритериального выбора компенсирующих мероприятий

Если сценарий крупного возмущения $v \in V$ разбит на T^v периодов, то общая постановка задачи функционирования ТЭК при воздействии крупного возмущения v в период времени $t=1, \dots, T^v$ может быть описана в следующем виде:

$$\sum_{v \in V} \sum_{t=1}^{T^v} [(cx_t^v + bs_t^v) + p(R_t - y_t^v) + hw] \rightarrow \min \quad (2.1)$$

$$s_{t-1}^v + A^v x_t^v + Uq_t^v - y_t^v - s_t^v \geq 0, \quad (2.2)$$

$$x_t^v \leq D_t^v, \quad (2.3)$$

$$y_t^v \leq R_t \quad (2.4)$$

$$q_t^v \leq wQ_t, \quad (2.5)$$

$$s_t^v \leq S_t^v, \quad (2.6)$$

$$s_0^v = S_0, \quad (2.7)$$

где x_t^v – искомый вектор, элементы которого характеризуют интенсивность применения технологических способов функционирования элементов СЭ (добычи, переработки,

преобразования и транспорта энергоресурсов); y_t^v – искомый вектор, элементы которого характеризуют объёмы потребления отдельных видов энергоресурсов; q_t^v – искомый вектор, описывающий интенсивность проведения мероприятий по повышению адаптивности; s_t^v – искомый вектор, компоненты которого характеризуют объёмы запасов топлива; A^v – матрица, описывающая технологии производства и передачи энергоресурсов, значения элементов которой зависят от воздействия возмущения v ; D_t^v – вектор, определяющий технически возможные интенсивности применения отдельных технологических и производственных способов, значения элементов которого зависят от воздействия возмущения v ; R_t – вектор, элементы которого демонстрируют потребности в отдельных видах энергоресурсов; U – матрица, отражающая локализацию проведения мероприятий по повышению адаптивности; Q_t – вектор, задающий пределы интенсивности проведения мероприятий по повышению адаптивности; S_t^v – вектор, определяющий ёмкость хранилищ, значения элементов которого зависят от воздействия возмущения v ; c – вектор, элементы которого определяют удельные затраты по каждому технологическому способу функционирования элементов СЭ; b – вектор удельных затрат на эксплуатацию хранилищ; p – вектор удельных ущербов, возникающих вследствие недопоставки отдельных видов энергоресурсов потребителям; h – вектор, задающий удельные затраты на подготовку и проведение мероприятий по повышению адаптивности.

Целевая функция (2.1) является свёрткой трёх критериев. Первый критерий отражает издержки, связанные с функционированием ТЭК. Второй критерий оценивает ущерб от дефицита энергоресурсов из-за воздействия возмущения v . Затраты на подготовку и проведение мероприятий по повышению адаптивности характеризуются третьим критерием.

Воздействие возмущения v реализуется элементами матрицы A^v и векторов D_t^v , S_t^v в уравнениях (2.2), (2.3) и (2.6) соответственно. Их элементы характеризуют степень деформации различных компонентов СЭ вследствие воздействия возмущения в период времени t .

Уровень необходимого снабжения потребителей отдельными видами энергоресурсов задаётся уравнением (2.4). Технические ограничения на проведение мероприятий по повышению адаптивности определяется в (2.5).

Объём запасов топлива в хранилищах в период времени t ограничен их доступной ёмкостью согласно неравенству (2.6). Уравнение (2.7) предполагает, что в начале шага сценария возмущения v (в период времени $t=0$) все хранилища имеют некоторый начальный запас энергоресурсов, описываемый вектором S_0 .

3. Апробация методики выбора компенсирующих мероприятий в исследованиях живучести ТЭК страны

Апробация заявленной методики выполняется на примере задачи реконструкции Единой системы газоснабжения (ЕСГ) России [3-4]. Апробируется многопериодная оптимизационная модель функционирования ТЭК в постановке (2.1 - 2.7).

Эта модель представлена тремя топливными отраслями (угольной, газовой, нефтеперерабатывающей), электро- и теплоэнергетикой. Объекты всех отраслей представлены на агрегированном уровне с суммарными значениями технологических показателей по территории. Газовая отрасль в модели представлена реальной расчетной схемой ЕСГ России с включением таких объектов, как магистральные газопроводы (486 объектов), компрессорные станции (24 объекта), подземные хранилища данных (24 объекта), источников газа (28 объектов). Потребители топлива в модели представлены объектами электро- и теплоэнергетики, прочими внутренними потребностями, странами-экспортерами. Потребители вторичных энергоресурсов (электроэнергии и тепла) – суммарными потребностями территорий, экспортерами. Элементами учета живучести в модели являются запасы топлива на начало расчетного периода, диверсификация топлива (замена газа на топочный мазут и обратно) на тепловых электростанциях и в котельных. Временной аспект исследований – сутки, территориальный аспект – субъекты РФ.

В рамках апробации для условий январской пиковой нагрузки моделируется отключение критически-важных объектов газовой отрасли (элементов, чья неработоспособность способна привести к серьезным сбоям поставок энергоресурсов). Устранение последствий этих отключений реализуется путем реконструкции сети ЕСГ. Определяются новые участки магистральных газопроводов между узлами сети, прежде не связанные между собой. В качестве критериев отбора вводимых элементов сети используется ограничение по длине участков сети. Элементы логической взаимозависимостей КМ пока не рассматриваются. Рассматриваются отключения, приводящие к существенным дефицитам тепла по стране в целом.

4. Заключение

Предложена методика однокритериального выбора набора инвариантных КМ. Методика строится на оптимизации многопериодной модели функционирования ТЭЖ в рамках одного ВЭ. Данная модель позволяет по временным интервалам моделировать комбинации каждой НС с полным набором типовых КМ, что обеспечивает инвариантность мероприятий. Оптимальность выбора мероприятий гарантируется конкуренцией штрафов за недопоставку энергоресурсов с суммарными капитальными вложениями на ввод КМ по всем временным интервалам. Методика применима для исследований с относительно небольшой расчетной частью, ее безусловным преимуществом является возможность получения уже готового набора инвариантных мероприятий в рамках рассматриваемых НС. Основным недостатком методики является невозможность задания критерия выбора мероприятий, «защитого» в целевой функции модели.

5. Благодарности

Исследование выполнено в рамках проекта государственного задания № FWEU-2021-0003 (рег. номер: АААА-А21-121012090014-5).

6. Список источников

- [1] Методы и модели исследования живучести систем энергетики / Антонов Г. Н., Черкесов Г. Н., Криворуцкий Л. Д. и др. Новосибирск: Наука. Сиб. Отд-ние, 1990. 285 с.
- [2] Еделев А.В., Береснева Н.М., Костромин Р.О. Методика выбора алгоритмов оптимизации живучести энергетических инфраструктур // Системы анализа и обработки данных. 2023. № 4 (92). С. 97-129.
- [3] Воробьев С. В., Еделев А. В., Сендеров С. М. Проблемы поиска критически важных объектов газовой отрасли с позиций энергетической безопасности России // Управление развитием крупномасштабных систем MLSD'2018. 2018. С. 44-49.
- [4] Надежность топливо- и энергоснабжения потребителей с позиции обеспечения энергетической безопасности / С.М. Сендеров [и др.]; отв. ред. Н.И. Воропай; Рос. Акад. Наук, Сиб. Отд-ние, Ин-т систем энергетики им. Л. А. Мелентьева. – Новосибирск: СО РАН, 2022. 132 с.
- [5] Сендеров С.М., Крупенёв Д.С., Воробьев С.В., Береснева Н.М., Бояркин Д.А. Методический подход к оценке уровня значимости критически важных объектов энергетики при взаимосвязанной работе энергетических отраслей // Известия Российской академии наук. Энергетика. 2023. № 2. С. 31-45.

Разработка генератора платформенно-зависимой модели для системы проектирования виртуальных помощников

Никита Допиро¹

¹ Иркутский национальный исследовательский технический университет, ул. Лермонтова 83, г. Иркутск, 664033, Россия

Аннотация

В данной статье будет рассмотрена разработка генератора платформенно-зависимой модели для системы проектирования виртуальных помощников на базе JUST AI с использованием платформы JAICP с помощью MDA подхода.

Ключевые слова

MDA, виртуальный помощник, чат-бот, JAICP, искусственный интеллект

1. Введение

Искусственный интеллект уже давно стал неотъемлемой частью нашей повседневной жизни. Одним из наиболее популярных применений ИИ являются виртуальные помощники – это сервис, способный взаимодействовать с людьми на основе уже имеющегося сценария, выполнять различные задачи и предоставлять необходимую информацию. В большинстве современных устройствах уже есть голосовой помощник, способный предоставить информацию о погоде, вызвать такси, заказать еду.

Алгоритм работы виртуального помощника [1], [2] включает в себя следующие этапы: принятие и распознавание голосового ввода; преобразование его в текстовый формат; обработка уже имеющегося текста на естественном языке; определение намерений пользователя; выявление задачи, ее параметров и определение дальнейших шагов решения; выполнение этапов решения выявленной задачи; создание ответа в виде диалога и представление результата пользователю.

Для реализации выше указанного алгоритма современные виртуальные помощники должны обладать определенным набором необходимых навыков: распознавать и понимать естественный язык (NLP), обеспечивающие распознавание и генерацию речи (голоса и текста); поддерживать многоканальную интеграцию; сохранять и использовать контекст диалога с пользователем; диалоговые элементы управления, относящиеся к системе управления диалогами; иметь доступ к БД и брать оттуда необходимую информацию для пользователя; компоненты для решения задач и принятия решений.

Актуальность проведения научных исследований в данном направлении подтверждается растущим спросом на использование виртуальных помощников (Сбер “Салют”, “Amazon Alexa”, Microsoft “Cortana”), а также увеличением разнообразия программных средств для их разработки. Применение виртуальных помощников широко распространено в компаниях и предприятиях для автоматизации клиентского обслуживания, и не только, которые в свою очередь направлены на оптимизацию производственной деятельности пользователя.

Данная работа вносит вклад в направления автоматизации разработки ВА в данной статье используется MDA подход. Model Driven Architecture (MDA) — модельно-ориентированный

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: nikitadopiro@gmail.com

ORCID:0009-0007-2621-9069



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.12

подход к разработке программного обеспечения. Принцип работы этого подхода состоит в создании абстрактной метамодели управления и обмена метаданными, то есть моделями. И в последующем задании способов ее преобразования в поддерживаемые технологии программирования, такие как C++, Java и другие. На примере [3], были сформулированы модели MDA и их трансформации (рис. 1).

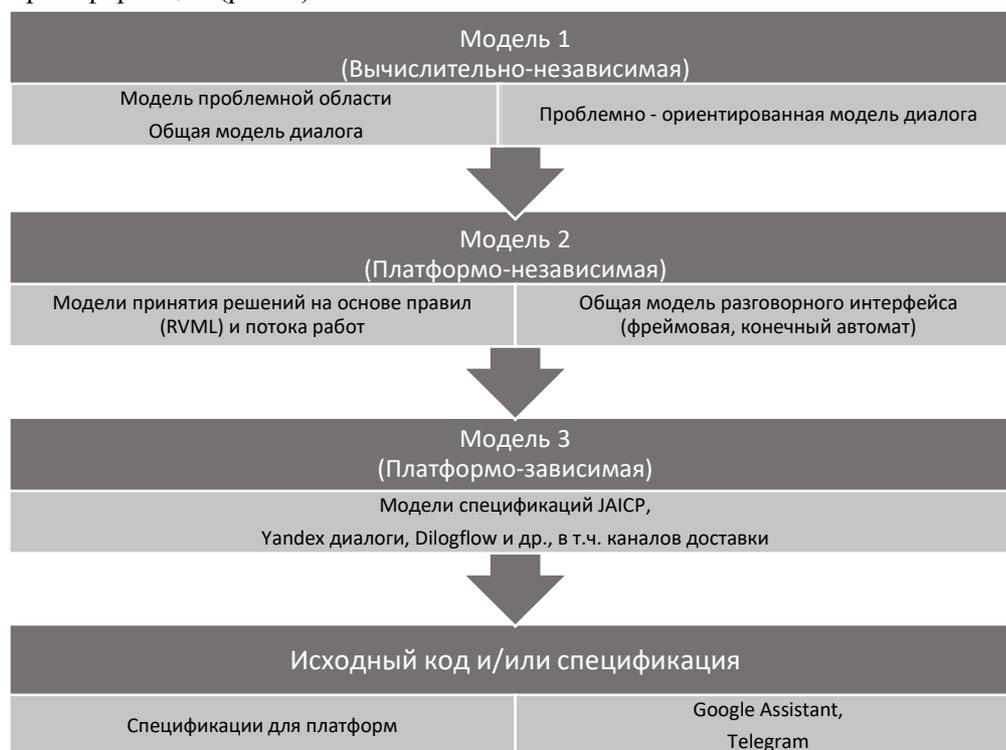


Рисунок 1: Концепция средства: модели MDA и их трансформации

2. Платформенно-зависимая модель JAICP

В качестве платформы разработки ВА выбрана JAICP. Just AI Conversational Platform (JAICP Ultimate) — это комплексное программное решение для разработки, внедрения, эксплуатации и поддержки интеллектуальных чат-ботов и голосовых помощников. Платформа использует передовые алгоритмы анализа естественного языка и эффективно сочетает подходы на основе правил и машинного обучения.

Согласно используемой MDA методологии была разработана платформенно-зависимые модели JAICP на языке Java. Далее будут рассмотрены основные классы и их взаимосвязи.

Диаграмма классов (рис. 2) описывает работу платформы JAICP, а также наглядно показывает его составляющие компоненты. Таким образом созданный ВА на данной платформе поддерживает многоканальность (Phone channels, Customer engagement platforms, Inbound channels), обработку естественного языка, благодаря встроенному NLU API (Caila, Dialogflow), что обеспечивает обработку сущностей, на базе платформы имеется два вида сущностей-пользовательские сущности (User entity) и системные сущности (System entity), а также поддерживает работу с переходами между состояниями, по принципу конечного автомата.

Состояние (рис. 3) может хранить контекст, структуру, которая представляет текущий контекст обработки запроса. JAICP также позволяет сохранять произвольные данные о текущей сессии. После выполнения реакций бота содержимое сессии сохраняется во внутренней базе данных и хранится между запросами пользователя. При завершении сессии эти данные очищаются. Переход между состояниями осуществляется благодаря тегам активации (н-р теги q или q!, intent или intent!) и реакций (н-р теги go или go!). JAICP позволяет гибко настроить выполнения тегов активации, разметить область действия (н-р параметр onlyThisState), или указать определённый переход в определённое состояние (н-р параметр ToState). Также в JAICP

имеется гибкая настройка обработки намерений, благодаря встроенному NLU API, при активации тегов intent или intent!

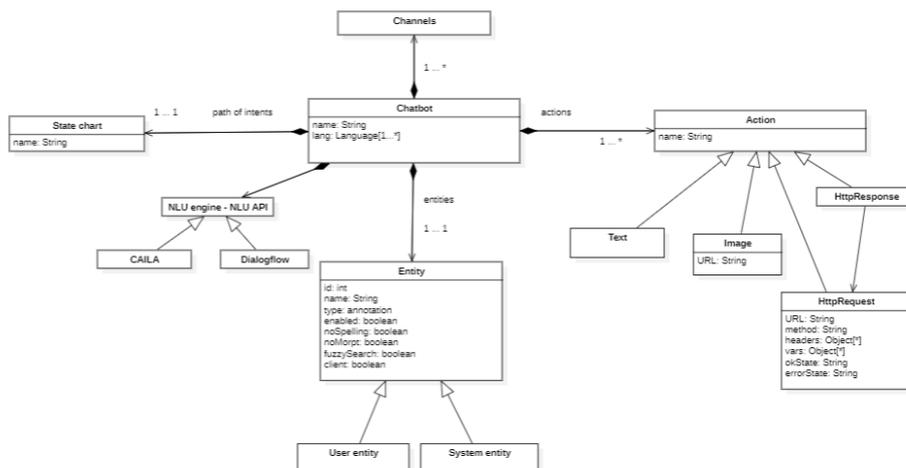


Рисунок 2: Диаграмма классов основных элементов JAICP

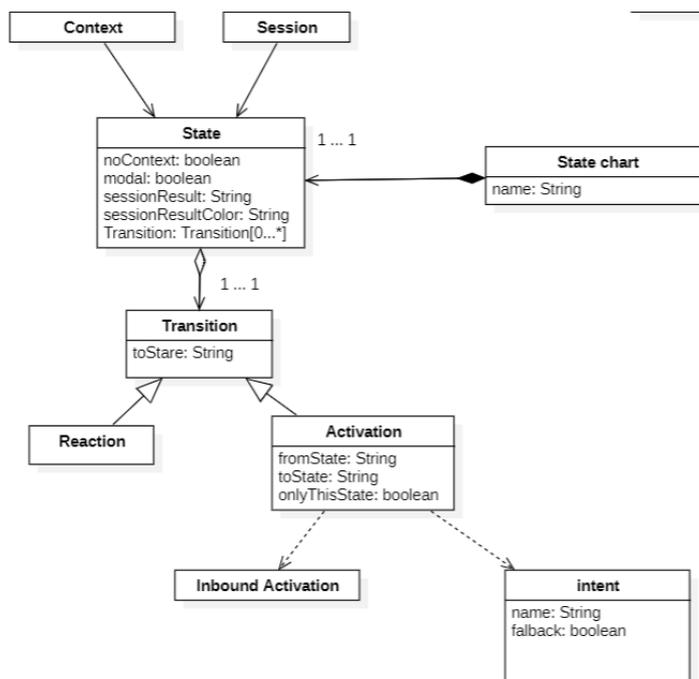


Рисунок 3: Диаграмма классов, описывающая взаимосвязь переходов и состояний

Диаграмма классов (рис. 4) описывает то, как обрабатываются намерения и сущности с заполнением слотов (слот-филлинг), и их компоненты. Обработка происходит следующим образом: фраза, которая была распознанная из общего контекста, благодаря заполненным паттернам и/или тренировочным фразам, представляет собой активацию для перехода в другое состояние, также в этом же контексте, где была распознана фраза, обрабатываются сущности, при распознавании которых заполняются определённые слоты. Заполненные слоты имеют следующие параметры: название — имя слота; сущность — системная, либо пользовательская сущность, задающая тип данных, которые попадут в слот; required — параметр активен, если слот обязателен для заполнения, чтобы обработать интент; массив — если параметр активен, то в слот помещаются все сущности данного типа, оформленные как массив; questions — указываются вопросы, которые задаются в процессе уточнения незаполненных слотов.

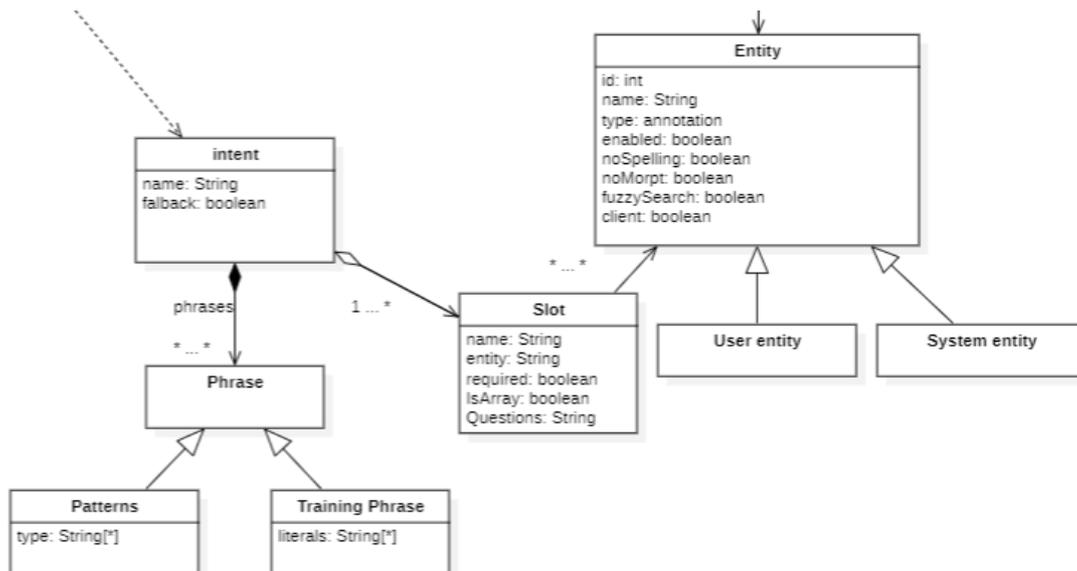


Рисунок 4.Классы для хранения информации о намерениях и сущностях с заполнением слотов, и их компонентов

3. Иллюстративный пример

В качестве примера был разработан чат-бот для поиска и ремонта неисправностей стиральной машины. Для формирования множества возможных неисправностей и соответствующего множества рекомендаций по их устранению было использовано руководство по ремонту неисправностей стиральных машин “Samsung” [4]. Разработанный чат-бот функционирует путем анализа описания обнаруженной неисправности в стиральной машине со стороны пользователя. В зависимости от характера и описания выявленной проблемы, бот переходит в различные состояния диалога. С учетом последующих ответов пользователя, чат-бот принимает решения относительно методов устранения указанной неисправности.

Чат-бот включает следующие состояния: старт, начальное состояние бота; меню; стиральная машинка не функционирует; в резервуар не поступает вода; не запрограммированная остановка во время стирки; не работает таймер; барабан не крутится; отключения бота; обработка исключений; и следующие сущности: стиральная машина; резервуар; барабан; таймер.

Сущности и намерения были описаны в терминах общей модели разговорного интерфейса [1]. Состояния и переходы между ними описаны в разрабатываемой в настоящее время проблемно-ориентированной модели диалога для диагностика неисправности бытовых приборов.

Задачей генератора является формирование спецификаций языкового интерфейса и кода управления диалогом на языке сценария JAICP DSL на основе разработанной и представленной выше платформенно-зависимой моделей JAICP:

Описание алгоритма работы генератора:

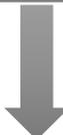
1. Заполняются объекты представленной информационной модели (интерфейса генератора) (фрагмент представлен на рис. 5).
2. Осуществляется генерация кода управления диалогом в форме языка сценария JAICP DSL на основе информации из объектов (фрагмент представлен на рис. 5).

```

String name_chatbot="chat_bot";
String lang_chatbot="ru";
Chatbot chatbot=new Chatbot(name_chatbot, lang_chatbot);
this.count_states++;

//State Start
States Start=new States( name_state "Start", session_code "Start", session_code "47E47D1", no_confirm false, mode false);
States.generationStates(Start);
Transaction q_teg_Start = new Transaction( transaction_name "reqact/start", no_confirm "");
Transaction.generation(q_teg_Start);
Action image_Start = new Action( url "https://www.gstatic.com/gsp/content/uploads/2019/02/83635615928.jpg");
Action.generation(image_Start);
Action a_Start = new Action( url "запустить! Вспомогатель (И по ремонту старых машин. Пожалуйста, оплатите или выберите категорию неисправности.");
Action.generation(a_Start);
//State The machine does not show any signs
States The_machine_does_not_show_any_signs=new States( name_state "The machine does not show any signs", session_code "The machine does not show any signs",
no_confirm false, mode false);
States.generationStates(The_machine_does_not_show_any_signs);
Transaction intent_the_machine_does_not_show_any_signs = new Transaction( transaction_name "The machine does not show any signs", no_confirm false,
no_confirm "The machine does not show any signs", no_confirm "");
Transaction.generation(intent_the_machine_does_not_show_any_signs);
Action a_the_machine_does_not_show_any_signs_1 = new Action( url "Чтобы провести устранение неисправности своими руками, нужно выполнить несколько наших рекомендаций.");
Action.generation(a_the_machine_does_not_show_any_signs_1);
Action a_the_machine_does_not_show_any_signs_2 = new Action( url "Вы готовы к их выполнению?");
Action.generation(a_the_machine_does_not_show_any_signs_2);

```



```

state: Incorrect setting of the washing machine
q: * (Нет/нет) * || fromState = "/The machine does not show any signs/Error code"
if: $session.the_machine_does_not_show_any_signs["incorrect setting of the washing machine"]==false
o: Возможно, вы выбрали не ту программу. Проверьте правильность настроек и запустите стиральную машинку.
o: Вы всё ещё наблюдаете неисправность?
script:
  $session.the_machine_does_not_show_any_signs["incorrect setting of the washing machine"]=true
else:
o: К сожалению у нас нет больше рекомендаций для исправления неисправности.
  Обратитесь в авторизованный сервисный центр для диагностики работы и ремонта техники.
go: /Bye

state: Incorrect setting not confirm
q: * (Да/да) * || fromState = "/The machine does not show any signs/Incorrect setting of the washing machine"
go: /The machine does not show any signs/Error code

state: Incorrect setting confirm
q: * (Нет/нет) * || fromState = "/The machine does not show any signs/Incorrect setting of the washing machine"
o: Отлично! Вам удалось устранить неисправность.
go: /Bye

state: Error code confirmation
q: * (Да/да) * || fromState = "/The machine does not show any signs/Error code"
if: $session_does_not_show_any_signs_Error_codes["dC3"]==false ||
  $session_does_not_show_any_signs_Error_codes["dDC"]==false ||
  $session_does_not_show_any_signs_Error_codes["DE_DE1_DE2_DC_DC1_DC2_ED"]==false
o: Ошибке код ошибки
script:
  $session.the_machine_does_not_show_any_signs["Error code"]=true;
else:
o: К сожалению у нас нет больше рекомендаций для исправления неисправности.
  Обратитесь в авторизованный сервисный центр для диагностики работы и ремонта техники.
go: /Bye

state: Confirm_dC3
intent: /dC3 || fromState = "/The machine does not show any signs/Error code confirmation"
if: $session_does_not_show_any_signs_Error_codes["dC3"]==true
go: /The machine does not show any signs/repair_end

script:
  $session_does_not_show_any_signs_Error_codes["dC3"]=true;
go: /The machine does not show any signs/The loading door is not closed

state: Confirm_dDC
intent: /dDC || fromState = "/The machine does not show any signs/Error code confirmation"
if: $session_does_not_show_any_signs_Error_codes["dDC"]==true
go: /The machine does not show any signs/repair_end

script:
  $session_does_not_show_any_signs_Error_codes["dDC"]=true;
go: /The machine does not show any signs/The loading door is not closed

```

Рисунок 5: Фрагмент входных и выходных данных работы генератора

4. Заключение

В представленной статье был проведен анализ и разработка генератора платформенно-зависимой модели для системы проектирования виртуальных помощников, основанных на технологии JUST AI с использованием платформы JAICP посредством Model-Driven Architecture (MDA) подхода. Основной целью исследования являлось создание эффективного инструмента для автоматизации процесса создания виртуальных помощников на данной платформе.

Были рассмотрены модели MDA и их трансформации, что позволило выявить основные преимущества данного подхода при разработке программного обеспечения.

Особое внимание было уделено объектным платформенно-зависимым моделям, используемым в рамках платформы JAICP. Подробно была описана структура этих моделей, их взаимосвязи и функциональные возможности, что способствовало более полному пониманию принципов работы данной системы.

В заключение статьи был представлен иллюстративный пример работы генератора платформенно-зависимой модели, с практическим применением разработанных концепций. Данный пример продемонстрировал эффективность и функциональность генератора по созданию виртуальных помощников на основе JUST AI с использованием платформы JAICP.

5. Список литературы

- [1] А.Б. Столбов, Н.О. Дородных, О.А. Николайчук, А.Ю. Юрин, Н.В. Допиро. Генерация и трансформация специализированных моделей для проблемно-ориентированного интеллектуального помощника //Материалы конференции «Ляпуновские чтения 2023». – Иркутск: ИДСТУ СО РАН, 2023. – 184 с.
- [2] A. Yurin, O. Nikolaychuk, N. Dorodnykh, A. Stolbov Towards knowledge-based virtual assistant development with the aid of ontology // International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). – 2022. – 830-834 pp.
- [3] Pérez-Soler S., Guerra E., de Lara J. Model-Driven Chatbot Development.Lecture Notes in Computer Science. – 2020. Vol. 12400, pp. 207–222.
- [4] Руководство по работе со стиральными машинами Samsung, 2024. URL: <https://www.samsung.com/ru/support/category/home-appliances/laundry/>

Система визуального моделирования знаний Knowledge Modeling System

Никита Дородных¹ и Александр Юрин¹

¹Институт динамики систем и теории управления имени В.М. Матросова СО РАН, ул. Лермонтова, 134, Иркутск, 664033, Россия

Аннотация

В работе описывается веб-ориентированная программная система – Knowledge Modeling System (KMS), обеспечивающая возможность визуального моделирования знаний в форме деревьев событий и диаграмм переходов состояний, а также генерации на их основе кодов баз знаний. Полученные таким способом базы знаний могут быть использованы в качестве прототипов. Приведено детальное описание разработанной системы, включая: метод создания баз знаний на основе преобразования концептуальных моделей; основные функции; архитектуру. Представлен иллюстративный пример использования системы KMS в контексте создания фрагмента базы знаний для решения задач диагностирования и прогнозирования технического состояния объектов в нефтехимии.

Ключевые слова

Визуальное моделирование, знания, база знаний, генерация кода, правила, онтология, дерево событий, диаграмма переходов состояний, трансформация моделей

1. Введение

Прикладные интеллектуальные системы поддержки принятия решений представляют собой сложные программные комплексы, создаваемые с целью повышения эффективности решения слабоформализованных задач. В архитектуре подобных систем выделяют базы знаний (БЗ) [1], содержащие информацию о закономерностях предметной области, а также решатели, реализующие алгоритмы логического вывода. Для представления знаний используются различные формализмы и языки, при этом формализмы, основанные на семантических технологиях (такие как онтологий [2] и графы знаний [3]) и логических правилах (продукциях) [4] сохраняют свою популярность. Однако, разработка как онтологических, так и продукционных БЗ продолжает оставаться одним из самых сложных и трудоемких этапов при создании интеллектуальных систем, который требует значительных ресурсов и привлечения специалистов различной квалификации.

Для решения данной проблемы в последние годы активно развиваются методы автоматического извлечения знаний из различных информационных источников (баз данных, веб-ресурсов, электронных документов и таблиц) [5]. Большинство предлагаемых решений в области разработки БЗ требуют от пользователей владения не только информацией предметной области, но знаний и навыков в области программирования. При этом актуально более полное вовлечение в данный процесс экспертов предметной области и системных

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: nikidomy@icc.ru (A. 1); iskander@icc.ru (A. 2)

ORCID: 0000-0001-7794-4462 (A. 1); 0000-0001-9089-5730 (A. 2)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.13

аналитиков за счет реализации методов автоматизированного преобразования данных из различных источников информации и визуального программирования.

Для поддержки визуального программирования и моделирования предлагаются различные редакторы построения когнитивных или концептуальных (информационных) моделей, в частности, концепт-карт (ИММС CMapTools, Miro, MindMaster), деревьев событий и отказов (ПО «Дерево неисправностей», Edraw, EFTed), диаграмм переходов состояний (StarUML, Enterprise Architect, Draw.io), диаграмм рыбе́й кисти (XMind, Fishbone Diagram Maker – SmartDraw, Creately) и др. Однако большинство из них не предназначены для целенаправленной разработки БЗ и не предоставляют какой-либо возможности для генерации кодов на определенном языке представления знаний (ЯПЗ) или интеграции построенных визуальных моделей БЗ определенной платформы. Данная особенность затрудняет практическое использование результатов моделирования иначе как в качестве «статических» графических артефактов и схем.

Данная работа посвящена описанию веб-ориентированной программной системы визуального моделирования знаний – Knowledge Modeling Language (KMS) [6], обеспечивающей возможность графического описания причинно-следственных зависимостей в форме деревьев событий и диаграмм переходов состояний и их отображения в коды БЗ на определенном ЯПЗ. При этом в качестве целевых ЯПЗ используются CLIPS (C Language Integrated Production System) и OWL (Web Ontology Language), а также формат электронных таблиц CSV (Comma-Separated Values). Разработанная система была использована для создания прототипов БЗ в области техногенной безопасности при решении задач диагностирования и прогнозирования технического состояния объектов и систем [7].

2. Программная система

Knowledge Modeling System (KMS) [6] – это веб-ориентированная программная система, предназначенная для моделирования знаний некоторой предметной области в форме концептуальных моделей и создания на их основе прототипов БЗ.

KMS реализует принципы технологии PEsOT (Prototyping Expert System based On Transformations) [8], которая рассматривает процесс создания интеллектуальных систем и БЗ в контексте последовательного преобразования моделей разного уровня абстракции [9].

2.1. Реализуемый метод разработки баз знаний

Метод разработки БЗ на основе трансформации концептуальных моделей, реализованный в KMS, образует последовательность из шести основных этапов:

Этап 1 – посвящен построению концептуальных моделей, описывающих некоторую предметную область. На данном этапе KMS обеспечивает поддержку построения:

1) Деревьев событий (event trees) [10], представляющих в виде иерархических древовидных структур сценарии развития событий. Деревья событий обычно используются для анализа и оценки рисков в различных областях, таких как природная и техногенная безопасность, управление проектами и др.

2) Диаграмм переходов состояний (state transition diagrams) [11], визуализирующих последовательность состояний и переходов между ними в виде ориентированного графа, где узлы – это состояния системы, а дуги (переходы) – это события или действия, которые приводят к изменению состояния системы. При этом могут быть указаны конкретные условия, выполняемые при каждом переходе. В отличие от деревьев событий данные диаграммы могут содержать циклы и петли. Диаграммы переходов состояний активно используются в проектировании и анализе сложных технических и социальных систем.

Этап 2 – представление (сериализация) построенных концептуальных моделей в формате XML (eXtensible Markup Language) с использованием специально разработанных спецификаций.

Этап 3 – анализ XML-структур представления концептуальных моделей с целью выделения элементов и их отношений.

Этап 4 – формирование модели продукций (правил) или онтологии по результатам анализа XML-структур. Выбор модели определяется целевым ЯПЗ, для которого планируется получение кода БЗ.

Этап 5 – генерация кода БЗ для целевой платформы путем преобразования сформированной модели продукций или онтологии.

Этап 6 – уточнение (модификация) полученных структур правил или онтологии.

В процессе перехода с этапа на этап происходит преобразование моделей, которые можно наглядно представить в виде таблицы соответствий (Таблица 1).

Таблица 1

Пример основных соответствий между элементами деревьев событий, модели продукций и конструкций на ЯПЗ CLIPS

Элемент формата представления деревьев событий	Элемент модели продукций	Оператор ЯПЗ CLIPS
TreeDiagram	Model	-
Node (name)	FactTemplate / Fact	Deftemplate
Parameter (name)	Slot	(slot "<name>")
Sequence	RuleTemplate / Rule	defrule
Node (parent_node)	FactTemplate / Condition	deftemplate / defrule (consequent)
Node (id)	FactTemplate / Action	deftemplate / defrule (antecedent)

Полученный таким образом код БЗ может быть использован в сторонних программных средствах (например, Personal Knowledge Base Designer (PKBD) [12] или Protégé) с целью уточнения и модификации и дальнейшего использования при разработке интеллектуальных систем.

2.2. Основные функции

Основные функции KMS:

- Создание, редактирование, просмотр и удаление пользователей (для администратора).
- Доступ к личному пространству пользователя (авторизация и аутентификация).
- Создание, редактирование, просмотр и удаление проектов в личном пространстве пользователя.
- Создание, редактирование, просмотр и удаление концептуальных моделей.
- Проверка корректности построенных диаграмм.
- Импорт и экспорт концептуальных моделей в виде сериализованных файлов в формате XML.
- Импорт онтологий в формате OWL (RDF/XML).
- Предоставление программного интерфейса (REST API) для доступа к моделям.
- Экспорт (генерация) кодов для целевых платформ.

2.3. Архитектура и особенности реализации

Разработанная веб-ориентированная программная система KMS обладает клиент-серверной архитектурой и включает следующие основные модули:

- Подсистема построения концептуальных моделей, включающая два визуальных редактора:
 - a. редактор классических и расширенных диаграмм деревьев событий (Extended Event Tree Editor) [13];

4. Заключение

Актуальность проблемы автоматизации разработки интеллектуальных систем обусловлена, в том числе, необходимостью создания нового методологического и программного инструментария, обеспечивающего более полное вовлечение в данный процесс непрограммируемых пользователей, а также использование уже накопленной ранее информации, представленной, в частности, в форме визуальных концептуальных моделей.

В данной работе предлагается веб-ориентированная программная система KMS [9], которая позволяет экспертам визуально моделировать знания предметной области в форме диаграмм переходов состояний и деревьев событий. Построенные таким образом концептуальные модели являются основным источником информации для автоматического формирования кодов БЗ как продукционного, так и онтологического типа. Приведено детальное описание предлагаемой системы, ее функции и архитектура. Также рассмотрен иллюстративный пример использования системы KMS в рамках создания фрагмента БЗ для решения задач диагностирования и прогнозирования технического состояния систем в нефтехимии.

5. Благодарности

Работа выполнена в рамках госзадания Минобрнауки России по проекту «Методы и технологии облачной сервис-ориентированной цифровой платформы сбора, хранения и обработки больших объемов разноформатных междисциплинарных данных и знаний, основанные на применении искусственного интеллекта, модельно-управляемого подхода и машинного обучения» (№ гос. регистрации: 121030500071-2).

6. Список литературы

- [1] Т.А. Гаврилова, Д.В. Кудрявцев, Д.И. Муромцев, Инженерия знаний. Модели и методы, СПб., Лань, 2023.
- [2] E. Karabulut, S.F. Pileggi, P. Groth, V. Degeler, Ontologies in digital twins: A systematic literature review. *Future Generation Computer Systems* 153 (2024) 442–456.
- [3] S. Ji, S. Pan, E. Cambria, P. Marttinen, P.S. Yu, A Survey on Knowledge Graphs: Representation, Acquisition and Applications, *IEEE Transactions on Neural Networks and Learning Systems* 33 (2022) 494–514.
- [4] A. Nowak-Brzezińska, A. Wakulicz-Deja, Exploration of rule-based knowledge bases: A knowledge engineer's support, *Information Sciences* 485 (2019) 301–318.
- [5] J.L. Martinez-Rodriguez, A. Hogan, I. Lopez-Arevalo, Information Extraction meets the Semantic Web: A Survey, *Semantic Web* 11 (2020) 255–335.
- [6] Knowledge Modeling System (KMS), 2024. URL: <http://kms.knowledge-core.ru/>.
- [7] А.Ф. Берман, К.А. Кузнецов, О.А. Николайчук, А.И. Павлов, А.Ю. Юрин, Информационно-аналитическая поддержка экспертизы промышленной безопасности объектов химии, нефтехимии и нефтепереработки, *Химическое и нефтегазовое машиностроение* 8 (2018) 30–36.
- [8] A.Yu. Yurin, Technology for Prototyping Expert Systems Based on Transformations (PESoT): A Method, *CEUR Workshop Proceedings* 2677 (2020) 36–50.
- [9] K. Czarnecki, S. Helsen, Feature-based survey of model transformation approaches, *IBM Systems Journal* 45 (2006) 621–645.
- [10] ГОСТ Р 54142-2010 Менеджмент рисков. Руководство по применению организационных мер безопасности и оценки рисков. Методология построения универсального дерева событий, М., Стандартинформ, 2020.
- [11] Дж. Хопкрофт, Р. Мотвани, Дж. Ульман, Введение в теорию автоматов, языков и вычислений, М., Вильямс, 2008.
- [12] Personal Knowledge Base Designer (PKBD), 2024. URL: <http://www.knowledge-core.ru/index.php?p=pkbd&lan=ru>.

- [13] Н.О. Дородных, Д.Н. Шпаченко, Extended Event Tree Editor, Свидетельство о государственной регистрации программы для ЭВМ, М., Рег. № 2020660788 от 11.09.2020.
- [14] Н.О. Дородных, Д.Н. Шпаченко, А.Ю. Юрин, State Transition Diagram Editor. Свидетельство о государственной регистрации программы для ЭВМ, М., Рег. № 2022664023 от 22.07.2022.
- [15] KMS GitHub, 2024. URL: <https://github.com/Lab42-Team/kms>.

Разработка модели системы газоснабжения на основе открытых данных для исследования живучести

Алексей Еделев^{1,2}, Глеб Данилов¹, Ольга Башарина^{1,3}

¹ Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

² Институт систем энергетики им. Л.А. Мелентьева СО РАН, ул. Лермонтова, д. 130, 664033, Иркутск, Россия

³ Уральский государственный экономический университет, ул. 8 Марта, д. 62, 620144, г. Екатеринбург, Россия

Аннотация

В данной статье описывается процесс разработки модели системы газоснабжения Европы на основе открытых данных, полученных в Ольденбургском институте сетевых систем энергетики (Германия). Необходимость разработки новых моделей отраслевых систем энергетики в настоящее время диктуется тем, что большинство подходов в области исследований живучести систем энергетики фокусируются на рассмотрении только одной из двух составляющих живучести, адаптации системы к крупным возмущениям или восстановлении производительности системы после крупного возмущения. Такая односторонность исследований приводит к выработке неэффективных стратегий по повышению живучести систем энергетики. Новые модели систем энергетики позволят более реалистично и точно моделировать функционирование этих систем в экстремальных условиях на всём интервале выживания, а не только во время их адаптации или восстановления. В настоящее время с помощью ряда инструментов и технологий ИДСТУ СО РАН построена современная схема газоснабжения Европы, а также осуществляется подготовка данных для расчёта потоков природного газа по построенной структуре системы газоснабжения.

Ключевые слова

Живучесть, модели систем энергетики, большие трубопроводные системы, открытые данные

1. Введение

Системы энергетики (СЭ) предназначены для добычи (производства, получения), переработки (преобразования), передачи (транспортирования), хранения и распределения энергоресурсов и снабжения ими потребителей [1]. Они относятся к критическим инфраструктурам, т.е. таким системам, нарушение функционирования которых отрицательно влияет на экономику страны и благополучие общества [2].

В основе функционирования и развития СЭ лежат требования надёжности, позволяющие данным системам справляться с уже известными возмущениями, вероятность возникновения которых высока, а по масштабу последствий их можно характеризовать как мелкие и средние. Однако, в настоящее время резко вырос интерес к исследованию способности СЭ выживать при столкновении с крупными возмущениями [3], вероятность возникновения которых весьма мала или достоверно неизвестна, а их воздействие может привести к тому, что СЭ без дополнительных мер не сможет далее выполнять возложенные на неё функции. При этом

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: flower@isem.irk.ru (A. 1); dangleb@isem.irk.ru (A. 2); basharinaolga@mail.com (A. 3)

ORCID: 0000-0003-2219-9754 (A. 1); 0000-0002-7151-782X (A. 3)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.14

изучается реакция СЭ на наступление таких экстремальных условий, последствия для потребителей, компенсация нежелательных последствий [4] и процесс восстановления системы [5].

Данная способность СЭ противостоять крупным возмущениям, не допуская их каскадного развития с массовым нарушением режима энергоснабжения потребителей, и восстанавливать исходное состояние объекта или близкое к нему, называется живучестью [1]. Главной целью исследования живучести является построение стратегий управления развитием и функционирования СЭ, которые позволяют эффективно адаптироваться к крупным возмущениям и восстанавливаться после них.

2. Проведение исследований живучести СЭ

Отправной точкой в современных исследованиях живучести СЭ являются источники (механизмы генерации) кривых живучести или временных рядов данных, представляющих состояния системы в последовательные моменты времени крупного возмущения [1, 6].

Говоря формально, состояние СЭ в момент времени $t = t_0, \dots, T$ описывается вектором x , в состав которого входят структурные и режимные параметры объектов энергетики, а также природно-климатические, социальные, экономические и прочие параметры, представляющие условия внешней среды. Если X это последовательность состояний СЭ $\{x^{t_0}, \dots, x^T\}$, то показатель производительности является функцией $P(t): X \rightarrow \mathbb{R}$, преобразующей x^t в скалярную величину (рис. 1). Поведение СЭ на определённом интервале времени характеризуется сводной метрикой $D: P \rightarrow \mathbb{R}$, которая сворачивает участок кривой живучести на рассматриваемом интервале времени в скалярную величину [6].

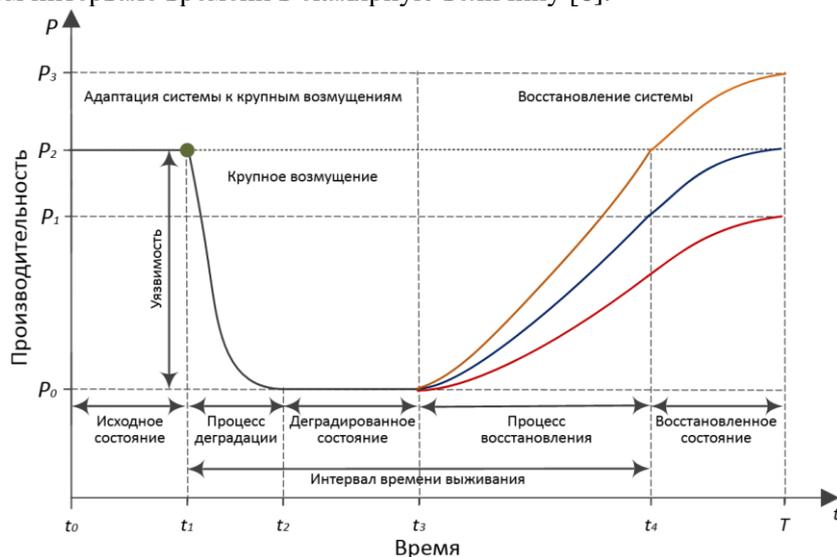


Рисунок 1: Идеализированная кривая живучести СЭ

Если показатели производительности отражают по большей части технологические или территориальные особенности СЭ, то сводные метрики характеризуют различные этапы поведения СЭ на интервале выживания (рис. 1).

Источники кривых живучести могут варьироваться в широких пределах: исторические данные, натурные опыты, агентное и имитационное моделирование системной динамики, искусственные нейронные сети, сетевые потоковые методы [7]. К последним относятся потоковые модели СЭ [21], которые можно записать в следующем общем виде:

$$\min_x C(X) \quad (1)$$

$$g_i(X) \leq 0, i = 1, \dots, I, \quad (2)$$

$$h_j(X) = 0, j = 1, \dots, J, \quad (3)$$

где X – искомая последовательность состояний СЭ, C – функция экономических затрат. Неравенства g_i в количестве I из (2) и равенства h_j в количестве J из (3) представляют разнообразные требования к функционированию и развитию СЭ.

Переход к оценке возможностей СЭ по адаптации к конкретному крупному возмущению \hat{v} происходит путём добавления в модель (1)-(3) вектора критериев живучести F :

$$\min_{X \in S(\hat{v})} R(X, \hat{v}) = (C(X, \hat{v}), F(X, \hat{v})) \quad (4)$$

где S – множество допустимых планов распределения потоков энергоресурсов в условиях, определяемых ограничениями (2), (3) и воздействием возмущения \hat{v} . Критерии F могут быть сводными метриками D для оценки живучести СЭ.

Параметризуя \hat{v} в (2), получается двухуровневая постановка для поиска наихудшего сценария возмущения из множества крупных возмущений V :

$$\max_{v \in V} \min_{X \in S(v)} R(X, v) \quad (5)$$

Аналогичным способом можно получить трехуровневую постановку для выбора наилучших мер по повышению живучести СЭ из множества W при реализации наихудшего возмущения из множества V :

$$\min_{w \in W} \max_{v \in V} \min_{X \in S(v, w)} R(X, v, w), \quad (6)$$

где S – множество допустимых планов распределения потоков энергоресурсов в условиях реализации возмущения $v \in V$ и меры по повышению живучести $w \in W$.

Вышеописанный процесс преобразования математической постановки (1) в (4) образует схему итеративного построения исследований живучести СЭ. Эта схема даёт возможность взглянуть на структуру исследований живучести глобально и правильно организовать вычислительные эксперименты, которые являются основным средством проведения исследований живучести СЭ в силу невозможности проведения натуральных опытов на них [1].

Организацией и проведением вычислительных экспериментов занимается предметно-ориентированная среда (ПОС), под которой понимается совокупность программно-аппаратных средств, позволяющих её конечным пользователям решать один или несколько классов задач некоторой предметной области. В данном случае ПОС учитывает особенности исследований живучести СЭ [8] и отображает представленную выше схему построения исследований живучести на вычислительные инфраструктуры с помощью разработанных в ИДСТУ СО РАН ряда инструментов и технологий их применения. В их число входят:

- распределённые пакеты прикладных программ (РППП), предназначенные для решения определённого класса задач [9];
- инструментарий ГеоАРМ [10] для создания автоматизированных рабочих мест, которые позволяют исследователю работать с исходными данными, включая слабоструктурированные, взаимодействовать с РППП, анализировать результаты вычислительных экспериментов.

3. Особенности моделирования отраслевых СЭ при исследовании их живучести

В основе архитектуры РППП для исследований живучести СЭ лежит обобщённый подход к моделированию взаимосвязанных технических систем, который основан на следующих принципах [11]:

- явное разделение моделирования любой системы на топологическую и функциональную составляющую;
- представление взаимосвязанных систем в виде метасистемы для моделирования их взаимодействия.

Согласно первому принципу, структурная модель отраслевой СЭ представляет собой сеть или направленный граф $G = (N, E)$, где N – это множество, состоящее из $n = |N|$ узлов, E – это множество, состоящее из $m = |E|$ дуг.

Множество узлов N делится на следующие подмножества: источники N_b , потребители N_d , промежуточные узлы N_o и хранилища N_c . Узлы могут быть выделенными реальными источниками, хранилищами или потребителями энергоресурсов. Однако чаще всего они представляют собой агрегированные группы объектов, которые имеют сходную

функциональность, достаточно однородны по своим характеристикам или располагаются в одном районе. Все узлы имеют географические координаты.

Для каждой дуги $(i, j) \in E$ узел $i \in N$ является началом, а узел $j \in N$ – концом. Также как в случае с узлами, дуги могут являться настоящими объектами транспорта, но в большинстве случаев они представляют собой агрегированные производственные возможности по передаче энергоресурсов.

Функциональная модель отраслевой СЭ описывает технологические цепочки от производства (добычи) до потребления отдельных видов энергоресурсов и может быть записана в виде следующей задачи линейного программирования [12]:

$$c^T x \rightarrow \min, \quad (7)$$

$$Ax = r, \quad (8)$$

$$\underline{x} \leq x \leq \bar{x}, \quad (9)$$

где c – вектор удельных затрат по технологическим способам функционирования действующих, реконструируемых или модернизируемых, а также вновь сооружаемых энергетических объектов; A – матрица технологических коэффициентов производства (добычи), переработки, транспорта, хранения и потребления энергоресурсов; x – искомый вектор, который, как следует из введения, характеризует интенсивность применения технологических способов функционирования энергетических объектов; \bar{x} – вектор, определяющий максимальные технически возможные интенсивности применения технологических способов функционирования энергетических объектов; \underline{x} – вектор, определяющий минимально технически требуемые интенсивности применения технологических способов функционирования энергетических объектов; r – вектор, задающий максимально возможные объёмы производства (отрицательные компоненты) и потребления энергоресурсов (положительные компоненты).

Говоря в целом, задача (7)–(9) находит поток минимальной стоимости по сети G . Если элемент a_{ij} матрицы A равен 1 для всех дуг $(i, j) \in E$ (каждый столбец матрицы A содержит только -1 и/или $+1$), то сеть G называется простой. Если a_{ij} не равен 1 хотя бы для одной дуги $(i, j) \in E$, то сеть G называется обобщенной.

Известно, что методы решения задач (7)–(9) для простых сетей [13] существенно более эффективны в вычислительном плане и, соответственно, намного быстрее, чем для обобщенных сетей. Если ранг матрицы A равен $n-1$, то любая обобщенная сеть может быть преобразована в простую с помощью специальных процедур масштабирования. В противном случае, если потери энергоресурсов при передаче довольно малы или удельные расходы при преобразовании близки к единице, то для повышения скорости решения потоковых задач множителями a_{ij} можно пренебречь, то есть считать их равными 1.

Последний способ используется в существующих моделях систем газоснабжения, нефте- и нефтепродуктоснабжения [14, 15], разработанных в ИСЭМ СО РАН для исследований живучести. Существенным недостатком данных моделей является их однопериодность, следствиями которой являются отсутствие динамики и упрощенное моделирование экстремальных условий функционирования СЭ.

Первое следствие однопериодности вышеуказанных моделей СЭ – статичность делает их пригодными только для исследования первой из сторон живучести: оценки системных возможностей по адаптации к крупным возмущениям, где одним из классов задач является анализ уязвимости [2]:

$$\max_{v \in V} \min_{X \in S(v)} \Lambda(X, v), \quad (10)$$

где метрика Λ отвечает за определение максимального падения производительности (рис. 1) вследствие воздействия крупного возмущения $v \in V$.

Второе следствие в виде упрощенного моделирования экстремальных условий функционирования СЭ, например неизменно постоянная пиковая потребность в энергоресурсах может привести к недопустимому завышению значений сводной метрики Λ [16].

Как результат, согласно работе [17] при равных экономических затратах стратегии, повышающие живучесть СЭ на всём интервале выживания, более эффективны чем стратегии, ориентированные только на снижение уязвимости.

Для решения перечисленных выше проблем, вызванных односторонностью исследований живучести, необходимы многопериодные потоковые модели, которые формируются как набор копий простой или обобщённой сети G , по одной на каждый момент времени $t = t_0, \dots, T$. Эти копии связываются между собой дугами, которые передают величину запасов энергоресурсов в хранилищах из одного периода времени в другой.

Новые функциональные модели систем газоснабжения, нефте- и нефтепродуктоснабжения позволят более реалистично и точно моделировать функционирование этих СЭ в экстремальных условиях. Например, падение давления газа вследствие воздействия крупного возмущения на газотранспортную сеть, может сказываться на ближайших к месту аварии потребителях через несколько часов, а на наиболее удалённых – через сутки и более. Также новые модели позволят исследовать зависимость системной производительности от непостоянства нагрузки и вариативности технических характеристик оборудования СЭ во время крупного возмущения и после него. Эта возможность является наиболее востребованной для трехуровневых задач оптимизации живучести (6), где внутри происходит оптимизация параметров работы СЭ в экстремальных условиях, задаваемых возмущениями $v \in V$.

4. Роль открытых данных в энергетических исследованиях

Открытость данных играет важную роль в разработке новых методов для решения задач (4)-(6), так как для сравнения их эффективности с другими подобными методами требуются четко определенные, реалистичные тестовые задачи. Примерами библиотек эталонных тестов, широко используемых в численной оптимизации, являются MINLPLib и MIPLIB. В сфере энергетики усилия в данном направлении направлены на разработку:

- программных платформ с открытым исходным кодом для оптимизации СЭ, например, Temoa с линейной эталонной моделью Utopia;
- открытых источников информации, содержащих массивы данных по энергетике страны или полные спецификации электростанций;
- общедоступных коллекций эталонных схем конкретных СЭ.

Если перейти к системам газоснабжения, то коллекция газотранспортных сетей GasLib широко применяется в различных исследованиях систем газоснабжения, включая их живучесть. В качестве известных открытых источников информации можно привести модель системы газоснабжения Бельгии. Результатом 3-хлетнего проекта SciGRID_gas, выполненного в Ольденбургском институте сетевых энергетических систем, стали выложенные в открытый доступ массивы данных по системе газоснабжения Европейского союза [76].

5. Создание модели системы газоснабжения Европы на основе данных проекта SciGRID_gas

В рамках проекта SciGRID_gas были получены три массива данных различной детализации по системе газоснабжения Европейского союза: IGGIELGNC-1, IGGIELGNC-2 и IGGIELGNC-3. Каждый из трёх массивов данных представляет собой набор текстовых файлов, содержащих список атрибутов объектов различных типов, и сами значения атрибутов. Также в рамках этого проекта были получены массивы данных SciGRID_gas CONS, которые содержат ежедневное потребление природного газа с 2010 по 2019 год с разбивкой на промышленный, жилой и коммерческий сектора для всех регионов стран Европейского союза.

В массивах данных IGGIELGNC-1, IGGIELGNC-2 и IGGIELGNC-3 выделяются четыре типа объектов: объекты добычи, объекты магистрального транспорта, объекты подземного хранения и объекты потребления. К объектам добычи можно условно отнести все объекты, выдающие газ в сеть магистрального транспорта, в том числе месторождения, терминалы сжиженного природного газа (СПГ) и подземные хранилища газа (ПХГ), если рассматриваемый момент времени совпадает со временем отбора газа из ПХГ. К объектам магистрального транспорта газа относятся участки магистральных газопроводов (МГ), включающие линейную часть и компрессорные станции (КС), расположенные на ней, а также пограничные пункты для

измерения объёмов газа, перемещаемого между странами. К объектам потребления продукции относятся группы потребителей, осуществляющих отбор газа из магистральных газопроводов в точках подключения к ним распределительных газопроводов, электростанции, работающие на газе, и ПХГ, если рассматриваемый момент совпадает с периодом закачки газа в ПХГ. Группы потребителей газа включают домашние хозяйства, промышленные предприятия и коммерческие объекты.

Согласно принципам подхода к моделированию СЭ, описанным выше, модель системы газоснабжения Европы на основе данных проекта SciGRID_gas делится на две части: топологическую и функциональную. Разработка модели СЭ начинается с топологической составляющей. Слабоструктурированные данные, которые представляют собой массивы IGGIELGNC-1, IGGIELGNC-2 и IGGIELGNC-3, были преобразованы авторами в соответствующие реляционные БД с помощью инструментария ГеоАРМ. Этот же инструментариий предоставляет возможности по редактированию структурной модели СЭ (рис. 2).

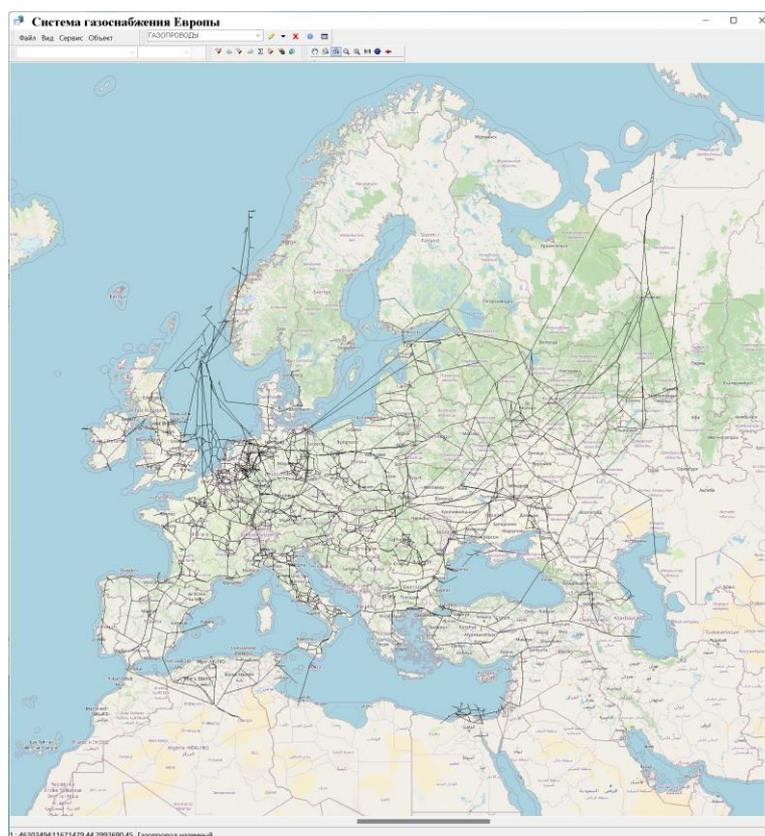


Рисунок 2: Структура системы газоснабжения Европы

В настоящее время авторами проводится верификация и валидация данных, необходимых для построения функциональной модели СЭ. Причиной проведения этих процедур является то, что многие значения атрибутов объектов ЭС в проекте SciGRID_gas были получены оценочным путём. Например, точные направления потока на большинстве участков МГ остаются неизвестными. Более того, некоторые участки МГ способны работать в двух направлениях. Поэтому, авторами проекта SciGRID_gas был разработан эвристический метод определения направления потока газа.

Необходимость верификации и валидации данных также подтверждается работой [19], в которой массивы данных SciGRID_gas использовались для создания газотранспортной сети Германии. В [19] сказано, что во время очистки данных были найдены и удалены 10 изолированных подсетей, которые не были связаны с основной сетью. Также было удалено примерно 50 участков МГ, у которых один и тот же узел был указан как входной и выходной, а длина этих участков МГ была установлена нулевой.

Валидация данных из массива SciGRID_gas CONS по Германии в работе [19] была проведена путём их сопоставления с подобными данными из БД DemandRegio. Результаты сравнения

показали, что для промышленного сектора DemandRegio, как правило, устанавливает более высокие значения спроса на природный газ, чем SciGRID_gas CONS. Что касается жилого и коммерческого секторов, то значения в этих массивах данных практически совпадают.

В ходе верификации и валидации данных выяснилось, что решатель задач линейного программирования Ip_solve, использующийся в существующих моделях отраслевых СЭ, разработанных в ИСЭМ СО РАН для исследований живучести, плохо справляется с размерностью модели газоснабжения Европы. Поэтому, авторами был выполнен переход на современные пакеты математической оптимизации SCIP [20] и HiGHS [21].

6. Заключение

В данной статье описывается процесс разработки модели системы газоснабжения Европы на основе открытых данных, полученных в Ольденбургском институте сетевых энергетических систем (Германия). Необходимость разработки новых моделей отраслевых СЭ в настоящее время диктуется тем, что большинство подходов в области исследований живучести СЭ фокусируются на рассмотрении только одной из двух составляющих живучести: адаптации системы к крупным возмущениям или восстановлении производительности системы после крупного возмущения. Такая односторонность исследований приводит к выработке неэффективных стратегий по повышению живучести СЭ. Новые модели СЭ позволят более реалистично и точно моделировать функционирование этих систем в экстремальных условиях на всём интервале выживания, а не только во время их адаптации или восстановления.

В настоящее время с помощью ряда инструментов и технологий ИДСТУ СО РАН построена современная схема газоснабжения Европы, а также осуществляется подготовка данных для расчёта потоков природного газа по построенной структуре системы газоснабжения.

7. Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 23-29-00460.

8. Список литературы

- [1] Надежность систем энергетики и их оборудования / Под общей редакцией Ю.Н. Руденко: В 4-х т. Т. 1: Справочник по общим моделям анализа и синтеза надежности систем энергетики / Под ред. Ю.Н. Руденко. М.: Энергоатомиздат, 1994. 480 с.
- [2] Zio E. Challenges in the vulnerability and risk analysis of critical infrastructures // Reliability Engineering and System Safety. 2016. Vol. 152. P.137-150.
- [3] Gasser P., Lustenberger P., Cinelli M., Kim W., Spada M., Burgherr P., Hirschberg S., Stojadinović B., Sun T. Y. A review on resilience assessment of energy systems // Sustainable and Resilient Infrastructure. 2019. P. 1–27.
- [4] Воропай Н. И., Криворуцкий Л.Д., Руденко Ю.Н., Шер И.А. Основные методические принципы исследования и обеспечения живучести систем энергетики // Антонов Г.Н., Черкесов Г.Н., Криворуцкий Л.Д. Методы и модели исследования живучести систем энергетики. Новосибирск: Наука. Сибирское отделение, 1990. 285 с.
- [5] Voropai, N., Rehtanz, C., 2019. Flexibility and Resiliency of Electric Power Systems: Analysis of Definitions and Content. In EPJ Web of Conferences (Vol. 217, p. 01018). EDP Sciences.
- [6] Poulin C.R., Kane M.B. Infrastructure resilience curves: Performance measures and summary metrics. Reliability Engineering & System Safety. 2021. Vol. 216. P. 107926. DOI: 10.1016/j.ress.2021.107926.
- [7] Иерархическое моделирование систем энергетики / под. ред. Н.И. Воропая, В.А. Стенникова. Новосибирск: Академическое изд-во «Гео», 2020. 314 с.
- [8] Еделев А.В., Карамов Д.Н., Башарина О.Ю. Методика оценки живучести автономной микросети // Информационные и математические технологии в науке и управлении. 2023. – № 4(32). С.117-126. DOI: 10.25729/ESI.2023.32.4.010.

- [9] Бычков И.В., Горский С.А., Еделев А.В., Костромин Р.О., Сидоров И.А., Феоктистов А.Г., Фереферов Е.С., Федоров Р.К. Поддержка управления живучестью систем энергетики на основе комбинаторного подхода // Известия РАН. Теория и системы управления. 2021. № 6. С. 122–135. DOI: 10.31857/S000233882106007X.
- [10] Фереферов Е.С., Бычков И.В., Хмельнов А.Е. Технология разработки приложений баз данных на основе декларативных спецификаций // Вычислительные технологии. 2014. Т. 19. № 5. С. 85-100.
- [11] Еделев А. В., Сендеров С. М., Береснева Н. М., Сидоров И. А., Феоктистов А. Г. Распределённая вычислительная среда для анализа уязвимости критических инфраструктур в энергетике // Системы управления, связи и безопасности. 2018. №3. С. 197-231.
- [12] Ahuja R K, Magnanti T L, Orlin J. B. Network flows: Theory, Algorithms, and Applications, 1993, Prentice-Hall, Inc.
- [13] P. Kovács. Minimum-cost flow algorithms: an experimental evaluation. Optimization Methods and Software, 2015, vol. 30, no. 1, pp. 94–127.
- [14] Храмов А.В., Еникеева С.М. и др. Программное и информационное обеспечение решения задач живучести Единой системы газоснабжения СССР // Методы и модели исследования живучести систем энергетики. Под ред. Руденко Ю.Н. Новосибирск: Наука, Сиб. отд. 1990. С.86-91
- [15] Клименко С.М., Сендеров С.М., Янченко В.А. Исследование проблем повышения устойчивости и экологической безопасности магистральных нефте- и нефтепродуктопроводов // Новые информационные технологии управления развитием и функционированием трубопроводных систем энергетики. Иркутск: СЭИ СО РАН, 1993. С. 119-127.
- [16] Zalitis, I., Dolgicers, A., Zemite, L., Ganter, S., Kopustinskas, V., Vamanu, B., Finger, J., Fuggini, C., Bode, I., Kozadajevs, J. and Häring, I., 2022. Mitigation of the impact of disturbances in gas transmission systems. International Journal of Critical Infrastructure Protection, 39, p.100569.
- [17] Ouyang M., Liu C., Xu M. Value of resilience-based solutions on critical infrastructure protection: Comparing with robustness-based solutions. Reliability Engineering & System Safety, 2019, 190, p.106506
- [18] Pluta, A., Medjroubi, W., Diettrich, J.C., Dasenbrock, J., Tetens, H.P., Sandoval, J.E. and Lünsdorf, O., 2022, April. SciGRID_gas-Data Model of the European Gas Transport Network. In 2022 Open Source Modelling and Simulation of Energy Systems (OSMSES). pp. 1-7.
- [19] Di Francesco, L., 2023. Modelling and Simulation of the German Gas Transmission Network with Hydrogen Blending (Doctoral dissertation, Politecnico di Torino).
- [20] Bestuzheva, K., Besançon, M., Chen, W.K., Chmiela, A., Donkiewicz, T., van Doornmalen, J., Eifler, L., Gaul, O., Gamrath, G., Gleixner, A. and Gottwald, L., 2023. Enabling research through the SCIP optimization suite 8.0. ACM Transactions on Mathematical Software, 49(2), pp.1-21.
- [21] Parallelizing the dual revised simplex method, Q. Huangfu and J. A. J. Hall, Mathematical Programming Computation, 10 (1), 119-142, 2018. DOI: 10.1007/s12532-017-0130-5.

Синтез микросетей для энергоснабжения сельскохозяйственных ферм Вьетнама

Алексей Еделев^{1,2}, Дмитрий Карамов^{1,3}, Van Doan Binh⁴, Nguyen Thi Thu Huong⁴

¹ Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

² Институт систем энергетики им. Л.А. Мелентьева СО РАН, ул. Лермонтова, д. 130, 664033, Иркутск, Россия

³ Иркутский национальный исследовательский технический университет, ул. Лермонтова, д. 83, 664074, Иркутск, Россия

⁴ Институт энергетики и окружающей среды Вьетнамской Академии наук и технологий, ул. Хоанг Куок Вьет, д. А9, 18, Ханой, Вьетнам

Аннотация

В данной статье приведены начальные результаты совместной работы авторов по задаче синтеза микросетей для сельскохозяйственных ферм Вьетнама. На основе проведённого анализа литературы было решено строить синтез микросетей в виде двух- или трёхуровневой структурно-параметрической оптимизации локальных энергетических комплексов. Для моделирования микросетей была выбрана концепция энергетического хаба.

Ключевые слова

микросети, структурно-параметрическая оптимизация, энергетический хаб, энергетический комплекс

1. Введение

Микросети являются разновидностью энергетических комплексов локального уровня, в которых совместное использование возобновляемых источников энергии с традиционными энергетическими установками и накопителями электрической энергии является экономически эффективным способом энергоснабжения потребителей. Особую значимость и большое распространение возобновляемые источники энергии получили при электрификации труднодоступных регионов мира, не имеющих связи с традиционными топливно-энергетическими комплексами [1].

Совместная работа накопителей электрической энергии в автономных системах электроснабжения имеет ряд факторов, влияющих как на техническую, так и экономическую эффективность использования тех или иных типов накопителей. В большей степени они касаются неравномерности генерации источниками, использующими энергию ветра и солнца в качестве первичных двигателей.

Накопители электрической энергии в микросетях, как правило, представлены разного рода аккумуляторными батареями. Они, в отличие от элементов генерации и трансформации электрической энергии, имеют определенный ряд специфических особенностей, которые необходимо учитывать при моделировании. Исходя из условий их эксплуатации [2], к данным особенностям можно отнести внутренние процессы старения активной массы аккумуляторных батарей, включая расслоение электролита, разрушение положительного электрода, тяжелую/необратимую сульфацию, кристаллизацию и т.п.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: flower@isem.irk.ru (A. 1); dmitriy.karamov@mail.ru (A. 2); binh.doan@ies.vn (A. 3); thuhuong.ies@gmail.com (A. 4)

ORCID: 0000-0003-2219-9754 (A. 1); 0000-0001-5360-4826 (A. 2); 0000-0001-5360-4826 (A. 3); 0000-0001-5360-4826 (A. 4)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.15

Стохастическая составляющая при генерировании электрической энергии существенно сказывается на основных эксплуатационных показателях аккумуляторных батарей. Например, в автономных фотоэлектрических системах эксплуатация аккумуляторных батарей сопровождается процессами с ярко выраженным циклическим характером в течение суток.

Способы образования микросетей можно разбить на следующие большие группы: преобразование существующих систем электроснабжения, динамическое формирование микросетей, синтез многоресурсных микросетей и интеллектуальных энергетических хабов [3].

2. Синтез многоресурсных микросетей

Синтез многоресурсных микросетей является задачей параметрической или структурно-параметрической оптимизации энергетических комплексов [4].

Параметрическая оптимизация отвечает за выбор типоразмеров и числа единиц технологического оборудования, определение оптимальных параметров работы технологического оборудования для каждого временного шага. Однако больший вклад в улучшение характеристик энергетических комплексов, снижение капитальных и эксплуатационных затрат вносит структурно-параметрическая оптимизация, отвечающая за подбор типов технологического оборудования и мест его установки.

Структурно-параметрическая оптимизация обычно формулируется поверх параметрической, в сравнении с которой является более сложной проблемой с математической и вычислительной точек зрения. Это выражается, в частности, в том, что пространство возможных топологических альтернатив для систем с такой высокой структурной и динамической сложностью, как ЭИ, может быть в принципе априори неизвестно [5]. Также математически доказано, что проблема структурно-параметрической оптимизации интегрированных систем энергетики является *NP*-трудной задачей [6].

Структурно-параметрическая оптимизация энергетических комплексов выполняется по двух- или трёхуровневой схеме, где на верхних уровнях происходит подбор типов технологического оборудования и мест его установки, выбор типоразмеров и числа единиц. Информация о выбранном оборудовании передаётся на нижний уровень, где производится оптимизация режимов работы энергетического комплекса с учётом решений, принятых на верхних уровнях. Затем расчётные данные об эффективности выбранного оборудования возвращаются нижним уровнем обратно наверх.

В литературе представлено множество подходов к структурно-параметрической оптимизации энергетических комплексов [7, 8].

3. Сложность моделей энергетических комплексов

Так как задача структурно-параметрической оптимизации энергетических комплексов в целом является *NP*-сложной, то добиться синтеза многоресурсных микросетей за приемлемое время достаточно сложно. Последнее обстоятельство влияет на вычислительные затраты, точность результатов расчётов и практическую применимость найденных конфигураций энергетических комплексов [9]. При этом, как говорят исследования [10-12], повышение детализации (сложности) энергетических моделей, влекущее за собой увеличение вычислительных затрат, не обязательно гарантирует рост точности моделирования или нахождение более эффективных системных конфигураций. По этим причинам в литературе до сих пор остаётся открытым вопрос подбора необходимой степени сложности моделей структурно-параметрической оптимизации энергетических комплексов.

После анализа нескольких десятков моделей структурно-параметрической оптимизации энергетических комплексов в работах [9, 11, 13] были выделены следующие факторы, определяющих сложность данных моделей:

- временная детализация исходных данных [14] и горизонт планирования [15],

- пространственная детализация исходных данных [16],
- тип модели (линейная, нелинейная и т.д.) [17] и подход к учёту неопределенности данных [8, 18],
- детализация представления и моделирования различных компонент энергетических комплексов [19].

Вышеперечисленные четыре фактора скорее определяют сложность модели энергетического комплекса нижнего уровня схемы структурно-параметрической оптимизации, нежели верхнего уровня. Сложность модели энергетического комплекса верхнего уровня по большей части определяется типом численного метода поиска оптимальных решений, используемым для оптимизации развития энергетических комплексов. Данные методы можно разделить на две дополняющие друг друга группы: эвристические и оптимизационные [20]. Эвристические методы, как правило, эффективны для быстрого получения хороших решений. Однако эти методы не могут гарантировать оптимальности, главным образом из-за последовательного характера и математически недоказанной сходимости. Оптимизационные методы, учитывающие структурные особенности ЭИ одновременно с конструктивными требованиями способны найти точное решение для практически любой целевой функции [21]. Обратной стороной оптимизационных методов по сравнению с эвристическими методами являются высокие вычислительные затраты и потенциально долгое время счёта.

4. Энергетический хаб

Для моделирования микросетей в настоящее время часто используется идея энергетического хаба, предложенная в 2007 году [22]. Она описывает в обобщенном виде технологии преобразования и накопления энергии. Энергетический хаб представляет собой посредника между различными энергетическими инфраструктурами и потребителями. На входы хаба, которые соединены с внешними энергосистемами, подаются различные энергоресурсы, такие как электричество и различные виды топлива: природный газ, уголь и т.д. Выходы хаба соединяются с потребителями, которые обеспечиваются энергетическими услугами необходимого качества, такими как электроснабжение, отопление и хладоснабжение. В хабе энергия преобразуется и аккумулируется с использованием технологий комбинированного производства тепла и электроэнергии, трансформаторов, силовой электроники, компрессоров, теплообменников и прочего оборудования [23].

Энергетический хаб обычно состоит из следующих элементов: n входов, d преобразователей энергии, k накопителей энергии и m выходов [24]. На рисунке 1 показан пример энергетического хаба, состоящего из 2 входов: электроэнергии и топлива, 2 выходов: электроэнергии и теплоэнергии, 3 преобразователей: электрического трансформатора, мини ТЭЦ для производства тепла и электроэнергии, электрического котла для покрытия пиковых потребностей в тепле, 2 накопителей: электрического аккумулятора и накопителя тепла.

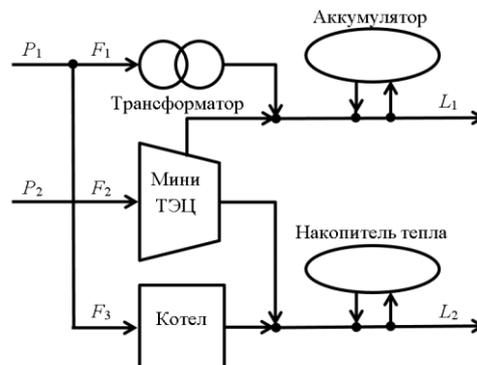


Рисунок 1: Пример энергетического хаба

Распределение потоков энергии от входов энергетического хаба к его выходам обычно описывается в матричной форме [25]:

$$\mathbf{L} = \boldsymbol{\eta}\mathbf{V}^{-1}\mathbf{P} - \mathbf{S}\mathbf{Q}^{ch} + \mathbf{S}\mathbf{Q}^{dis} + \mathbf{U}, \quad (1)$$

где \mathbf{P} – вектор, компоненты которого соответствуют потокам на входах хаба, \mathbf{L} – вектор, компоненты которого соответствуют потокам на выходах хаба, $\boldsymbol{\eta}$ – матрица преобразования, компоненты которой равны значениям коэффициентов преобразования одного вида энергии в другой, \mathbf{V} – матрица диспетчеризации потоков от входов хаба до преобразователей, \mathbf{S} – матрица связывания потоков в накопителях энергии с выходными потоками хаба, \mathbf{Q}^{ch} – вектор, компоненты которого соответствуют потокам заряда в накопителях энергии, \mathbf{Q}^{dis} – вектор, компоненты которого соответствуют потокам разряда в накопителях энергии, \mathbf{U} – вектор, компоненты которого равны значениям непокрытой части нагрузки на выходах хаба.

Потоки к конверторам можно выразить через входные потоки \mathbf{P} как

$$\mathbf{F} = \mathbf{V}^{-1}\mathbf{P}, \quad (2)$$

где \mathbf{F} – вектор, компоненты которого соответствуют потокам в преобразователях. С учетом (2) уравнение (1) можно переписать как

$$\mathbf{L} = \boldsymbol{\eta}\mathbf{F} - \mathbf{S}\mathbf{Q}^{ch} + \mathbf{S}\mathbf{Q}^{dis} + \mathbf{U}. \quad (3)$$

Для примера на рисунке 1 матрицы и вектора из уравнений (1)-(3) будут иметь следующий вид:

$$\mathbf{P} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_{11} & \eta_{21} & 0 \\ 0 & \eta_{22} & \eta_{32} \end{bmatrix}, \quad \mathbf{Q}^{ch} = \begin{bmatrix} Q_1^{ch} \\ Q_2^{ch} \end{bmatrix}, \quad \mathbf{Q}^{dis} = \begin{bmatrix} Q_1^{dis} \\ Q_2^{dis} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix},$$

где P_1 – входной поток электроэнергии, P_2 – входной поток топлива, L_1 – выходной поток электроэнергии, L_2 – выходной поток теплоэнергии, F_1 – поток электроэнергии через трансформатор, F_2 – поток природного газа в мини ТЭЦ, F_3 – поток электроэнергии в электрический котел, η_{11} – коэффициент полезного действия трансформатора, η_{21} – удельный расход топлива на выработку электроэнергии в мини ТЭЦ, η_{22} – удельный расход топлива на выработку тепла в мини ТЭЦ, η_{32} – коэффициент полезного действия электрического котла, Q_1^{ch} – поток заряда в электрический аккумулятор, Q_1^{dis} – поток разряда из электрического аккумулятора, Q_2^{ch} – поток теплоэнергии в накопитель, Q_2^{dis} – поток теплоэнергии из накопителя, U_1 – непокрытая часть потребности в электричестве, U_2 – непокрытая часть потребности в тепле.

5. Синтез микросетей для сельскохозяйственных ферм Вьетнама

На текущий момент авторами разработана типовая схема энергоснабжения сельскохозяйственных ферм Вьетнама, показанная в виде энергетического хаба на рисунке 2.

Микросеть на рисунке 2 состоит из следующих компонент: фотоэлектрические преобразователи (ФЭП), сетевой инвертор (СИН), аккумуляторные батареи (АБ), батарейный инвертор (БИН), ветроэлектрическая установка (ВЭУ) и установка для комбинированной выработки тепла и электричества из биомассы (миниТЭЦ). Также микросеть имеет подключение к внешней энергосистеме для покупки электроэнергии в случае нехватки собственных мощностей. Максимальная электрическая нагрузка не превышает 500 кВт.

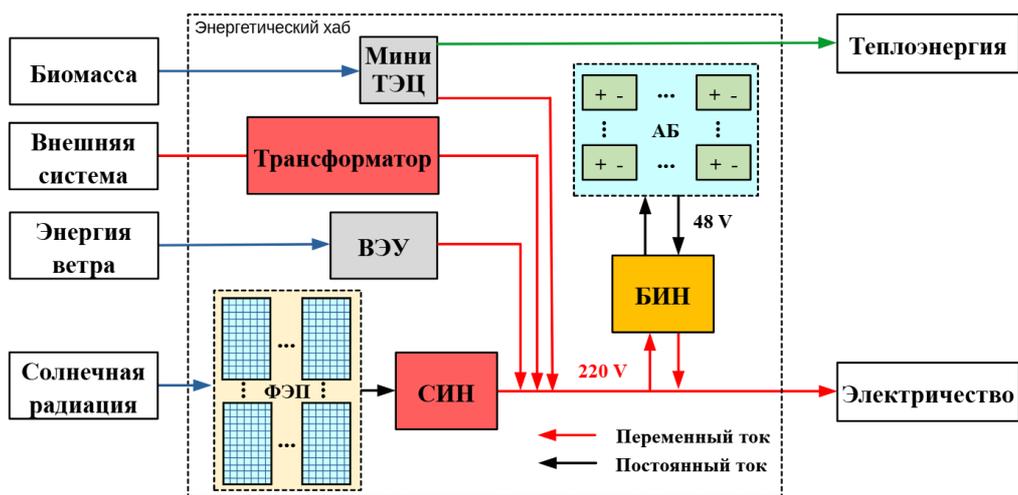


Рисунок 2: Пример энергетического хаба

Предполагается, что энергоснабжение сельскохозяйственных ферм контролируется цифровым двойником микросети, который подчиняется следующим правилам:

- если суммарная генерация микросети превышает уровень нагрузки, то выполняется прямое снабжение потребителя и аккумулярование энергии в АБ;
- если генерации от микросети недостаточно, то необходимый объем энергии поступает из АБ и внешней энергосистемы.

Синтез микросетей для сельскохозяйственных ферм Вьетнама будет заключаться в выборе типоразмеров и числа единиц оборудования для ФЭП, АБ, ВЭУ и миниТЭЦ.

6. Заключение

В данной статье приведены начальные результаты совместной работы авторов по задаче синтеза микросетей для сельскохозяйственных ферм Вьетнама. На основе проведенного анализа литературы было решено строить синтез микросетей в виде двух- или трёхуровневой структурно-параметрической оптимизации локальных энергетических комплексов. Для моделирования микросетей была выбрана концепция энергетического хаба.

7. Благодарности

Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды», , рег. № 124052100088-3).

8. Список литературы

- [1] Карамов Д.Н. Интеграция процесса категоризации электрохимических накопителей энергии в задачу оптимизации состава оборудования автономных энергетических комплексов, использующих возобновляемые источники энергии // Известия Томского политехнического университета. Инжиниринг георесурсов. 2019. Т. 330, № 5. С. 113–130.
- [2] Svoboda V., Wenzl H., Kaiser R., Jossen A., Baring-Gould I., Manwell J., Lundsager P., Bindner H., Cronin T., Nørgard P., Ruddell A., Perujo A., Douglas K., Rodrigues C., Joyce A., Tselepis S., van der Borg N., Nieuwenhout F., Wilmot N., Mattera F., Sauer D. U. Operating conditions of batteries in off-grid renewable energy systems. *Solar Energy*, 2007, 81, pp. 1409–1425.

- [3] Hussain, A., Bui, V.H. and Kim, H.M., 2019. Microgrids as a resilience resource and strategies used by microgrids for enhancing resilience. *Applied energy*, 240, pp.56-72.
- [4] Sustaining Power Resources through Energy Optimization and Engineering / P. Vasant and Nikolai Voropai, editors. Description: Hershey PA: Engineering Science Reference, 2016 by IGI Global. - 126 p.
- [5] Wang, L., Yang, Z., Sharma, S., Mian, A., Lin, T.E., Tsatsaronis, G., Maréchal, F. and Yang, Y. A review of evaluation, optimization and synthesis of energy systems: methodology and application to thermal power plants. *Energies*, 2018, vol. 12(1), p.73.
- [6] Goderbauer, S., Comis, M. and Willamowski, F.J., 2019. The synthesis problem of decentralized energy systems is strongly NP-hard. *Computers & Chemical Engineering*, 124, pp.343-349
- [7] Frangopoulos, C.A. Recent developments and trends in optimization of energy systems. *Energy*, 2018, vol. 164, pp. 1011-1020.
- [8] Schmeling, L., Schönfeldt, P., Klement, P., Vorspel, L., Hanke, B., von Maydell, K. and Agert, C. A generalised optimal design methodology for distributed energy systems. *Renewable Energy*, 2022, vol. 200, pp.1223-1239.
- [9] Cuisinier, E., Lemaire, P., Ruby, A., Bourasseau, C. and Penz, B., 2023. Impact of operational modelling choices on techno-economic modelling of local energy systems. *Energy*, 276, p.127599.
- [10] Nolting L, Praktijnjo A. Is the more complex model always the better one? Evidence from the assessment of security of electricity supply. In: *ENERDAY 2019 – the 13th international conference on energy economics and technology*. doi:10.13140/RG.2.2.32770.63684.
- [11] Wirtz, M., Hahn, M., Schreiber, T. and Müller, D., 2021. Design optimization of multi-energy systems using mixed-integer linear programming: Which model complexity and level of detail is sufficient?. *Energy Conversion and Management*, 240, p.114249.
- [12] Orth, R., Staudinger, M., Seneviratne, S.I., Seibert, J. and Zappa, M., 2015. Does model performance improve with complexity? A case study with three hydrological models. *Journal of Hydrology*, 523, pp.147-159.
- [13] Ridha, E., Nolting, L. and Praktijnjo, A., 2020. Complexity profiles: A large-scale review of energy system models in terms of complexity. *Energy Strategy Reviews*, 30, p.100515.
- [14] Hoffmann, M., Priesmann, J., Nolting, L., Praktijnjo, A., Kotzur, L. and Stolten, D., 2021. Typical periods or typical time steps? A multi-model analysis to determine the optimal temporal aggregation for energy system models. *Applied Energy*, 304, p.117825.
- [15] Teichgraber, H. and Brandt, A.R., 2022. Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities. *Renewable and Sustainable Energy Reviews*, 157, p.111984.
- [16] Simoes, S., Zeyringer, M., Mayr, D., Huld, T., Nijs, W. and Schmidt, J., 2017. Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria. *Renewable energy*, 105, pp.183-198.
- [17] Ommen, T., Markussen, W.B. and Elmegaard, B., 2014. Comparison of linear, mixed integer and non-linear programming methods in energy system dispatch modelling. *Energy*, 74, pp.109-118.
- [18] Gabrielli, P., Fürer, F., Mavromatidis, G. and Mazzotti, M., 2019. Robust and optimal design of multi-energy systems with seasonal storage through uncertainty analysis. *Applied energy*, 238, pp.1192-1210.
- [19] Gils, H.C., Pregger, T., Flachsbarth, F., Jentsch, M. and Dierstein, C., 2019. Comparison of spatially and temporally resolved energy system models with a focus on Germany's future power supply. *Applied Energy*, 255, p.113889.
- [20] Kumar, M., Soomro, A.M., Uddin, W. and Kumar, L. Optimal Multi-Objective Placement and Sizing of Distributed Generation in Distribution System: A Comprehensive Review // *Energies*. – 2022. – Vol. 5, N. 21. - P. 7850.
- [21] Wang, L., Yang, Z., Sharma, S., Mian, A., Lin, T.E., Tsatsaronis, G., Maréchal, F. and Yang, Y. A review of evaluation, optimization and synthesis of energy systems: methodology and application to thermal power plants // *Energies*. – 2018. – Vol. 12, N. 1. - P. 73.
- [22] Geidl M., Koepfel G., Favre-Perrod P., Klockl B., Andersson G., Frohlich K. Energy hubs for the future. *IEEE power and energy magazine*, 2007, 5(1), pp.24-30.
- [23] Xu, Y., Yan, C., Liu, H., Wang, J., Yang, Z. and Jiang, Y., 2020. Smart energy systems: A critical review on design and operation optimization. *Sustainable Cities and Society*, p.102369.

- [24] Huang, W., Zhang, N., Wang, Y., Capuder, T., Kuzle, I. and Kang, C., 2020. Matrix modeling of energy hub with variable energy efficiencies. *International Journal of Electrical Power & Energy Systems*, 119, p.105876.
- [25] Aghamohamadi, M. and Mahmoudi, A., 2019. From bidding strategy in smart grid toward integrated bidding strategy in smart multi-energy systems, an adaptive robust solution approach. *Energy*, 183, pp.75-91.

Структурно-параметрическая оптимизация энергоисточников в задачах развития систем теплоснабжения

Ольга Еделева¹, Алексей Еделев^{1,2}

¹Институт систем энергетики им. Л.А. Мелентьева СО РАН, ул. Лермонтова, д. 130, 664033, Иркутск, Россия

²Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

Аннотация

В статье рассмотрено решение задачи структурно-параметрической оптимизации энергоисточников систем теплоснабжения. Среди задач развития систем теплоснабжения она является наиболее актуальной для Байкальской природной территории (БПТ) из-за жестких экологических ограничений, накладываемых на работу энергоисточников. Разработана математическая модель оптимального развития энергоснабжающей системы с учетом представительных дней с часовой тепловой и электрической нагрузкой потребителя и учётом атмосферного загрязнения. Проведена апробация модели на практическом примере развития схемы энергообеспечения небольшого поселка, находящегося на территории БПТ. Подготовка исходных данных и решение задачи структурно-параметрической оптимизации энергоисточников осуществлялась с помощью сервисов цифровой платформы экологического мониторинга БПТ.

Ключевые слова

Система теплоснабжения, энергоисточник, тепловая сеть, структурно-параметрическая оптимизация, загрязнение атмосферы, представительные дни, график нагрузки, Байкальская природная территория, экологический мониторинг, цифровая платформа.

1. Введение

В рамках цифрового мониторинга Байкальской природной территории (БПТ) [1] разрабатываются новые модели, алгоритмы и прикладное ПО для моделирования природно-технических систем (ПТС) с упором на детализацию технологических процессов. ПТС представляют собой совокупность форм и состояний различных процессов взаимодействия компонентов природной среды с инженерными сооружениями на всех стадиях ее функционирования [2]. Необходимость разработки новых моделей, алгоритмов и ПО определяется интеллектуализацией управления ПТС, которое предъявляет повышенные требования к детализации моделируемых технологических процессов и качеству получаемых данных об условиях функционирования и развития ПТС. Одним из представителей ПТС являются системы энергетики, в частности, теплоснабжающие системы.

Проектирование систем теплоснабжения является задачей структурно-параметрической оптимизации и заключается в определении типов, мест расположения и состава оборудования энергоисточников в тесной взаимосвязи со схемой и параметрами тепловой сети (т.к. состав энергоисточников, их мощности и связи с потребителями определяют объем теплоносителя, передаваемого по тепловой сети). В случае расположения проектируемых энергоисточников на

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: edel@isem.irk.ru (A. 1); alexedelev@gmail.com (A. 2)

ORCID: 0000-0003-2219-9754 (A. 1); 0000-0003-2219-9754 (A. 2)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.16

территории БПТ в задачу структурно-параметрической оптимизации должна добавляться оценка их влияния на атмосферу.

Проектирование энергоисточников является многоэтапной процедурой, в данном исследовании делается допущение, что площадки для строительства определены в результате проведенного оптимизационного расчета тепловых сетей и находятся в местах скопления тепловых нагрузок. Следующим этапом является определение типов и типоразмеров оборудования энергоисточников. Тип энергоисточника определяется исходя из возможностей поставки топлива различных видов, и, кроме этого исследуется потребность в комбинированной выработке тепла и электроэнергии. Типоразмер имеет при этом важное значение, потому что, с одной стороны, с увеличением числа установок появляется возможность более мягкого маневрирования нагрузками, и повышается надежность энергоисточников. С другой стороны, бесконтрольное увеличение количества установок приводит к увеличению капитальных и эксплуатационных затрат. Согласно нормам проектирования [3], установленная тепловая мощность энергоисточника определяется по расчетной тепловой нагрузке. Представительные дни, в этом случае позволяют проверить, оправдана ли такая расчётная нагрузка. Как правило, представительные дни временного периода тепловой нагрузки, представляют собой выбор определенных дней из временного ряда, состоящего из нескольких лет, в которых замечены предельно высокие и низкие нагрузки, или же нагрузки имеют характерный, неравномерный характер в течение суток (как правило, это рабочие и выходные дни).

В [4] авторами был предложен один из возможных способов решения задачи структурно-параметрической оптимизации энергоисточников с помощью методологии энергетических хабов [5]. В данном исследовании этот способ был апробирован на примере небольшой энергоснабжающей системы с учетом представительных дней и оценкой атмосферного загрязнения.

2. Постановка задачи

В [6] была предложена двухуровневая модель выбора типа и состава оборудования энергоисточников (ЭИ) для решения задач перспективного развития энергоснабжающих систем различной размерности. На первом уровне модели происходит формирование множества структур ЭИ по их типу, на втором уровне подбирается оптимальный состав оборудования ЭИ с учетом надежности и рассчитываются затраты на сооружение и эксплуатацию микросети, включающие затраты в ЭИ и тепловую сеть (ТС). На этом же уровне оцениваются объемы выбросов загрязняющих веществ.

Основная идея применения методологии энергетических хабов в задачах развития систем теплоснабжения заключается в моделировании структуры варианта энергоисточника в виде энергетического хаба, $h = \{1, \dots, H\}$, который имеет множество входов M_h и множество выходов N_h . Каждый элемент из множеств M_h, N_h , связан с одним определённым энергоресурсом $k \in K$. Также хаб h имеет набор преобразователей в количестве J_h .

В хабе $h = \{1, \dots, H\}$ в период времени $t = \{1, \dots, T\}$ поток энергоресурса P_{hm}^t на входе $m \in M_h$ разделяется на потоки F_{hmj}^t к преобразователям $j = \{1, \dots, J_h\}$, связанными со входом $m \in M_h$

$$P_{hm}^t = \sum_{j \in \theta_h(m)} F_{hmj}^t, \quad (1)$$

$$F_{hmj}^t \leq \bar{F}_{hmj}, \quad (2)$$

где \bar{F}_{hmj}^t – пропускная способность связи преобразователя j со входом m хаба h , $\theta_h(m)$ – множество, содержащее преобразователи хаба h , связанные со входом m .

Образование потока энергоресурса L_{hn} на выходе $n \in N_h$ хаба $h = \{1, \dots, H\}$ показано на Рисунке 1. В период времени $t = \{1, \dots, T\}$ преобразователь энергии $j = \{1, \dots, J_h\}$ хаба h конвертирует поток F_{hmj}^t с коэффициентом преобразования η_{hmjn}^{con} и направляет его к выходу n . Поток L_{hn}^t на выходе $n \in N_h$ определяется как

$$L_{hn}^t = \sum_{m \in M_h} \sum_{j \in \theta_h(m,n)} \eta_{hmjn}^{con} F_{hmj}^t, \quad (3)$$

где $\theta_h(m, n)$ – множество, состоящее из преобразователей, посредством которых вход $m \in M_h$ соединяется с выходом n хаба h .



Рисунок 1: Переток энергии в энергетическом хабе от преобразователей к выходу

Коэффициент преобразования η_{hmjn}^{con} , помимо удельного расхода топлива, может также представлять удельные выбросы загрязняющих веществ в атмосферу.

Энергоисточники соединяются между собой и с потребителями с помощью сетей транспорта энергоресурсов (Рисунок 2). Каждая сеть представляет собой граф $G_k = (V_k, E_k)$, где V_k – множество узлов, E_k – множество ориентированных дуг, k – энергоресурс из множества K . На граф G_k могут накладываться дополнительные ограничения, связанные с физическими законами перемещения энергоресурса k .

В сеть G_k может входить множество узлов $V_k^g \subset V_k$, которое представляет обособленные источники энергоресурса $k \in K$, не относящиеся ни к одному из H хабов. Например, в V_k^g могут быть выделены источники топлива, которые на Рисунке 3.4 представлены прямоугольниками.

Через $V_k^d \subset V_k$ обозначается множество узлов потребления энергоресурса $k \in K$, которые на Рисунке 3.4 представлены кругами. Остальные узлы из множества $V_k^o = V_k \setminus (V_k^g \cup V_k^d)$ называются промежуточными и служат местами пересечения потоков энергоресурса в сети транспорта. Каждый узел v сети G_k должен удовлетворять одному из условий баланса потоков энергоресурса $k \in K$ в период времени $t = \{1, \dots, T\}$

$$g_{kv}^t + \sum_{(q,v) \in \Psi_k(v)} a_{kqv} f_{kqv}^t - \sum_{(v,q) \in \Omega_k(v)} f_{kvq}^t = 0, v \in V_k^g, \quad (4)$$

$$\sum_{(q,v) \in \Psi_k(v)} a_{kqv} f_{kqv}^t - \sum_{(v,q) \in \Omega_k(v)} f_{kvq}^t + u_{kv}^t = \bar{d}_{kv}^t, v \in V_k^d, \quad (5)$$

$$\sum_{(q,v) \in \Psi_k(v)} a_{kqv} f_{kqv}^t - \sum_{(v,q) \in \Omega_k(v)} f_{kvq}^t = 0, v \in V_k^o, \quad (6)$$

$$0 \leq u_{kv}^t \leq \bar{d}_{kv}^t, v \in V_k^d, \quad (7)$$

$$0 \leq f_{kvq}^t \leq \bar{f}_{kvq}, (v, q) \in E_k, \quad (8)$$

$$0 \leq g_{kv}^t \leq \bar{g}_{kv}^t, v \in V_k^g, \quad (9)$$

где $\Psi_k(v) \subseteq E_k$ – множество дуг, входящих в узел v , $\Omega_k(v) \subseteq E_k$ – множество дуг, исходящих из узла v , f_{kqv}^t – поток энергоресурса k по дуге $(q, v) \in E_k$ в период времени t , g_{kv}^t – производство энергоресурса k в узле $v \in V_k^g$ в период времени t , \bar{d}_{kv}^t – потребность узла $v \in V_k^d$ в энергоресурсе k в период времени t , u_{kv}^t – недопоставка энергоресурса k в узел $v \in V_k^d$ в период времени t , a_{kqv} – коэффициент, характеризующий потери при транспорте энергоресурса k по дуге $(q, v) \in E_k$, $\bar{f}_{kvq}, (v, q) \in E_k$ – пропускная способность дуги $(q, v) \in E_k$.

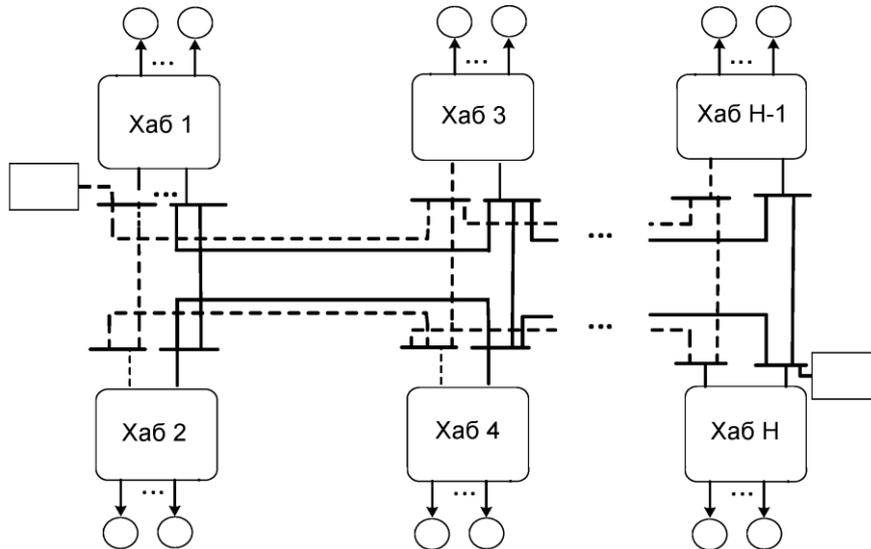


Рисунок 2: Сеть энергетических хабов

Если учитывать энергетические хабы, соединённые с G_k в узле v , то уравнения (4)-(6) для периода времени $t = \{1, \dots, T\}$ запишутся как

$$g_{kv}^t + \sum_{(q,v) \in \Psi_k(v)} a_{kqv} f_{kqv}^t + \sum_{h \in \Lambda_k^g(v)} L_{hn}^t - \sum_{h \in \Lambda_k^d(v)} P_{hm}^t - \sum_{(v,q) \in \Omega_k(v)} f_{kvq}^t = 0, v \in V_k^g, \quad (10)$$

$$\sum_{(q,v) \in \Psi_k(v)} a_{kqv} f_{kqv}^t + \sum_{h \in \Lambda_k^g(v)} L_{hn}^t - \sum_{h \in \Lambda_k^d(v)} P_{hm}^t - \sum_{(v,q) \in \Omega_k(v)} f_{kvq}^t + u_{kv}^t = \bar{d}_{kv}^t, v \in V_k^d, \quad (11)$$

$$\sum_{(q,v) \in \Psi_k(v)} a_{kqv} f_{kqv}^t + \sum_{h \in \Lambda_k^g(v)} L_{hn}^t - \sum_{h \in \Lambda_k^d(v)} P_{hm}^t - \sum_{(v,q) \in \Omega_k(v)} f_{kvq}^t = 0, v \in V_k^o, \quad (12)$$

где $\Lambda_k^d(v)$ – множество хабов, входы которых соединены с сетью G_k в узле v , P_{hk}^t – поток энергоресурса $k \in K$, поступающего на вход $m \in M_h$ хаба $h \in \Lambda_k^d(v)$ в период t , $\Lambda_k^g(v)$ – множество хабов, выходы которых соединены с сетью G_k в узле v , L_{hn}^t – поток энергоресурса $k \in K$, поступающего на выход $n \in N_h$ хаба $h \in \Lambda_k^g(v)$ в период t .

Целью оптимизации функционирования взаимосвязанных хабов является наиболее полное удовлетворение потребностей в энергоресурсах при минимальных затратах на эксплуатацию хабов и транспортных сетей на протяжении всех периодов времени $t = \{1, \dots, T\}$:

$$\sum_{t=1}^T \left[\sum_{h=1}^H \sum_{m \in M_h} C_{hk}^t P_{hm}^t Z_h^t + \sum_{k \in K} \sum_{v \in V_k^d} y_{kv} u_{kv}^t + \sum_{k \in K} \sum_{(v,q) \in E_k} c_{kvq} f_{kvq}^t \right] \rightarrow \min, \quad (13)$$

где C_{hk}^t – цена на энергоресурс $k \in K$, поступающего на вход $m \in M_h$ хаба $h = \{1, \dots, H\}$ в период t , P_{hm}^t – поток энергоресурса $k \in K$, поступающего на вход $m \in M_h$ хаба $h = \{1, \dots, H\}$ в период t , Z_h^t – булева переменная, определяющая, будет ли реализован хаб $h = \{1, \dots, H\}$, u_{kv}^t – недопоставка энергоресурса k в узел $v \in V_k^d$ в период времени t , y_{kv} – штраф за недопоставку энергоресурса $k \in K$ в узел $v \in V_k^d$ в период t , c_{kvq} – удельная стоимость транспорта энергоресурса $k \in K$ по дуге $(v, q) \in E_k$, f_{kvq}^t – поток энергоресурса $k \in K$ по дуге $(v, q) \in E_k$ в период t .

Ограничениями для целевой функции (13) выступают уравнения (7)-(12).

3. Практический пример развития схемы энергоснабжения поселка

Продемонстрируем пример развития схемы энергоснабжения небольшого поселка с населением в 6 000 чел., находящегося на территории БПТ в Восточной Сибири. Планируемая перспективная тепловая нагрузка поселка составит 23,19 Гкал/ч, электрическая – 102,91 тыс. МВт·ч в год, что связано с перспективной застройкой (рост численности населения) и необходимостью реконструкции объектов энергоснабжения. В настоящее время в поселке функционирует одна центральная котельная на буром угле (ЭИ №1, Рисунок 3).

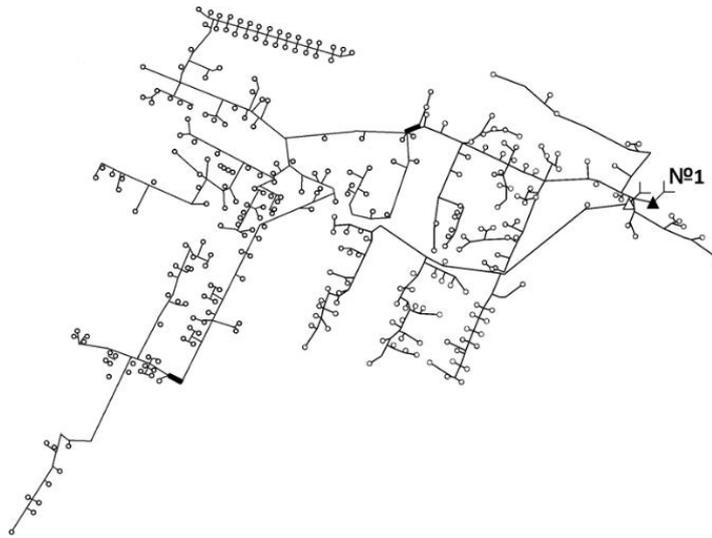


Рисунок 3: Пространственное представление конфигурации ТС с 1-им ЭИ для поселка

При формировании технологических вариантов ЭИ посёлка рассматривались два варианта топливоснабжения: подача газа на цели энергоснабжения (в рамках планируемой к реализации в регионе программы газификации) и использование щепы в качестве топлива (близкое расположение лесоперерабатывающего комбината по производству сухих пиломатериалов).

Был рассмотрен ввод 4 типов ЭИ: котельной (на газе и щепе) и мини ТЭЦ (на газе и щепе). При вводе котельных электроэнергия для поселка покупается из внешней энергосистемы, затраты на нее учтены в эксплуатационных расходах, в случае мини ТЭЦ варианты типоразмеров оборудования были подобраны с учетом полного электроснабжения поселка, при этом излишки электроэнергии заданы в качестве прибыли от ее продажи во внешнюю энергосистему.

Каждый вариант структуры ЭИ включает три типоразмера основного оборудования и задается следующим кодом: тип установки – типоразмер установки – количество установок – вариант конфигурации ТС. Для каждого варианта структуры ЭИ с учетом всех технико-экономических характеристик типоразмеров оборудования и физико-химического состава топлива были рассчитаны капиталовложения, эксплуатационные расходы и выбросы вредных веществ (твердые частицы, окись серы, окись азота, двуокись углерода, бенз(а)пирен), выбор оборудования был произведен на расчетную тепловую нагрузку с учетом нормативных документов [3].

Для задачи (7)-(13) были построены часовые графики тепловых (Рисунок 4) и электрических нагрузок (Рисунок 5), характеризующие представительные дни различных сезонов года. Представительные дни выбирались на основе анализа природно-климатических данных для места расположения для за период 10 лет. Выбор представительных дней осуществлялся с помощью сервисов получения и анализа временных рядов. Решение задачи (7)-(13) происходило с помощью сервиса структурно-параметрическая оптимизация ПТС. Результаты оптимизации приведены на Рисунке 6, где красным шрифтом выделены оптимальные варианты. Оба вышеуказанных сервиса входят в состав цифровой платформы экологического мониторинга БПТ [1].

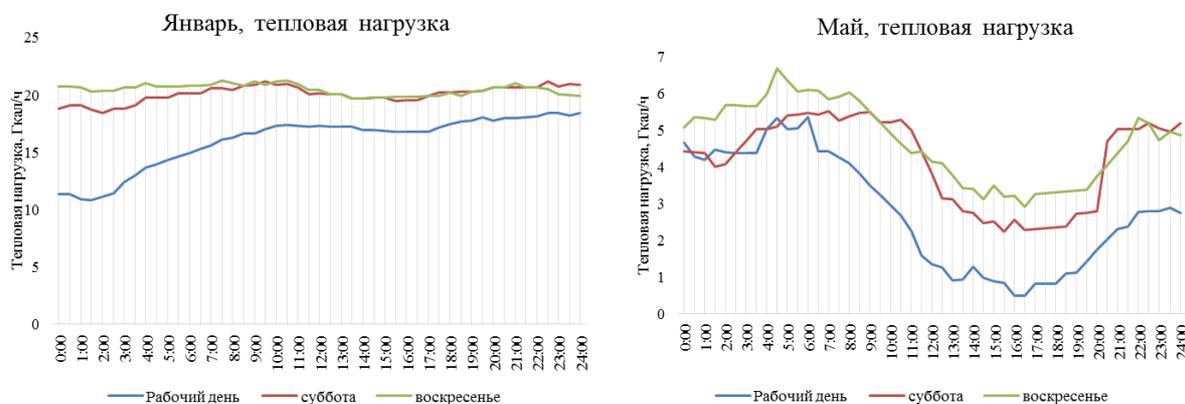


Рисунок 4: Графики часовой тепловой нагрузки представительных дней января и мая

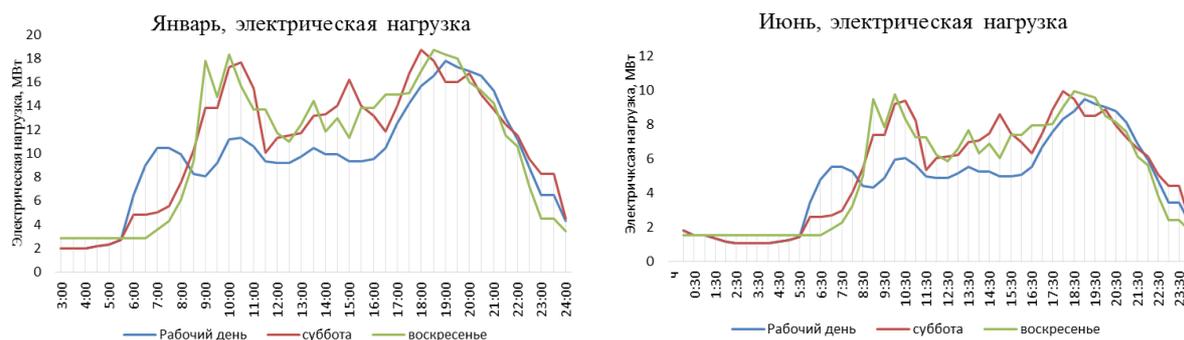


Рисунок 5: Графики часовой электрической нагрузки представительных дней января и июня

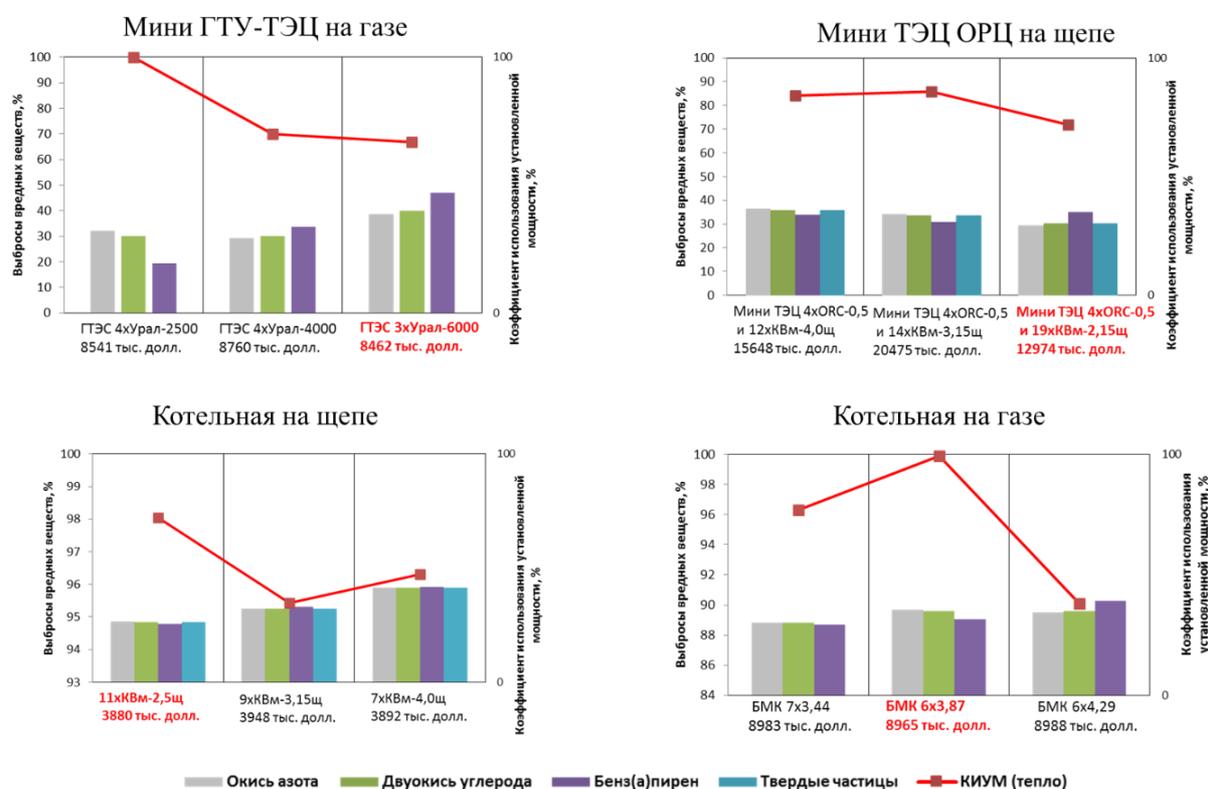


Рисунок 6: Результаты структурно-параметрической оптимизации структур ЭИ

Если сравнить оптимальные варианты между собой (Рисунок 7), можно выделить высокую степень технической эффективности и сравнительно недорогую стоимость вариантов котельных на щепе и на газе с покупкой электроэнергии из внешней энергосистемы. Варианты комбинированной выработки тепла и электроэнергии проигрывают вследствие увеличения мощности ЭИ для выработки электроэнергии (учитывалось полное покрытие потребности в перспективной электроэнергетической нагрузке поселка) и возрастания в связи с этим валовых выбросов вредных веществ, а также увеличения затрат. Кроме этого, анализируя коэффициент использования установленной мощности (КИУМ), выбранные типоразмеры оборудования по комбинированным источникам не будут обеспечивать оптимальную загрузку и эффективную работу оборудования ЭИ в течение года.

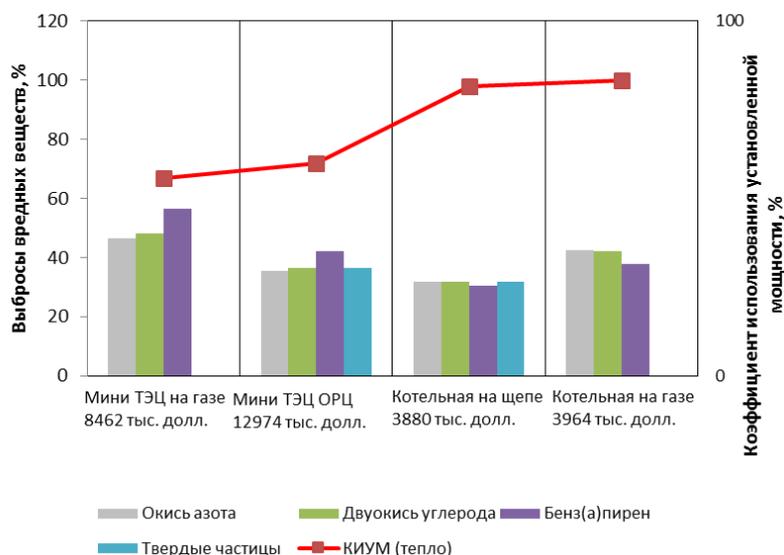


Рисунок 7: Сравнение оптимальных вариантов ЭИ

4. Выводы

Системы теплоснабжения являются одним из представителей ПТС, расположенных на территории БПТ. Традиционные энергоисточники систем теплоснабжения, сжигающие топливо для выработки тепла и электричества, относятся к экологически опасным объектам, вследствие чего необходимо оценивать влияние развития данных ПТС на экологию БПТ. Сделать это можно различными путями. В данной статье рассматривается внедрение учёта объёмов выбросов вредных веществ в постановку задачи структурно-параметрической оптимизации энергоисточников.

Предлагаемый в статье подход апробирован на примере развития схемы энергоснабжения небольшого поселка, находящегося на территории БПТ. Рассмотрено 4 варианта энергоснабжения, каждый из которых включает три типоразмера основного оборудования энергоисточников. Анализ вариантов производился на основе сравнения коэффициентов использования установленной мощности и объёмов выбросов вредных веществ, таких как твердые частицы, окись азота, двуокись углерода, бенз(а)пирен. Подготовка исходных данных и решение задачи структурно-параметрической оптимизации осуществлялась с помощью сервисов цифровой платформы экологического мониторинга БПТ.

5. Благодарности

Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды», рег. № 124052100088-3).

6. Список литературы

- [1] Цифровой мониторинг Байкальской природной территории. URL: <https://baikal-project.icc.ru/>.
- [2] Т. К. Салина, Концептуальный подход к управлению устойчивым развитием ТЭК Северо-Западного федерального округа, Журнал экономической теории, 2016, № 4, с. 80–91.

- [3] Справочник по проектированию котельных установок систем централизованного теплоснабжения: (общие вопросы проектирования и основное оборудование). – М.: Энергия, 1979. – 224 с.
- [4] V. A. Stennikov, O. A. Edeleva, E. A. Barakhtenko and D. V. Sokolov Methodological Approach to the Integrated Optimization of the Heat-Source Structure in the Problems of Developing Heat-Supply Systems // Thermal Engineering. – 2020. Vol. 67, No. 12, pp. 935-946. DOI: 10.1134/S0040601520120083
- [5] Almassalkhi, M. Optimization framework for the analysis of large-scale networks of energy hubs / M. Almassalkhi, I. Hiskens // 17th Power System Computation Conference (Stockholm, Sweden, August 22-26 2011). – 2011. – 7 p.
- [6] Edeleva O., Mikheev A. and Stennikov A. An Approach for Efficient Heat Source Selection in Long-Term Expansion Planning of Urban District Heating Systems // Mathematical Models and Methods of the Analysis and Optimal Synthesis of the Developing Pipeline and Hydraulic Systems, 2022. 3S Web Conf., Volume 397, 2023, DOI 10.1051/e3sconf/202339702002

Математическое моделирование солнечной радиации с учетом динамики в средних и нижних слоях атмосферы

Дмитрий Карамов¹, Евгений Юмашев¹

¹ Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

Аннотация

В статье представлена универсальная методика моделирования солнечной радиации на рассматриваемой территории с учетом динамики процессов, происходящих в средних и нижних слоях атмосферы. За основу была взята модель для условий ясного неба - Iqbal (1980) с последующей компиляцией с моделью учета облачности *F. Kasten - G. Czeplak* (1980). Показаны основные алгоритмы расчета, ключевые параметры и их связь с актинометрическими показателями.

В настоящий момент команда ИДСТУ СО РАН разрабатывает открытый WPS-сервис, позволяющий автоматически рассчитывать солнечную радиацию на рассматриваемой территории на основании ретроспективных данных формата FM 12 Synop.

Ключевые слова

Солнечная радиация, актинометрические показатели, WPS-сервис, Python, FM 12 Synop, METAR

1. Введение

Описание ключевых метеорологических параметров является первым шагом при использовании возобновляемых источников энергии. Метеорологические параметры позволяют качественно оценить потенциал того или иного возобновляемого энергоресурса, например, падающая на м² за год солнечная радиация, средняя за рассматриваемый промежуток времени скорость ветра с учетом повторяемости.

Метеопараметры могут быть представлены численными значениями и качественной (словесной) оценкой. Например, скорость ветра, давление, влажность, суммарная солнечная радиация, альbedo облаков и Земли, толщина озонового слоя и другие параметры представляются численными значениями. Качественная оценка характеризует отдельные метеоданные такие как состав облаков каждого яруса, наличие облаков плохой погоды, туманов, задымлений и т.д. Зачастую качественная оценка метеоданных связана с описанием ключевых параметров облачности, прозрачности нижних слоев атмосферы.

С позиции практического использования метеоданные могут быть разделены на две группы, а именно непосредственно влияющие на ключевые параметры работы генераторов использующих возобновляемую энергию, а также косвенные влияние которых учитывается через отдельные параметры и факторы.

Потенциал возобновляемых источников энергии характеризуется количеством суммарной солнечной радиации (кВт·ч/м²·год) и средней скоростью ветра за рассматриваемый период (месяц, год) в м/с. Солнечная радиация и скорость ветра необходимы для определения потенциально возможной генерации, моделирования отдельных режимных параметров и как результат анализа технической и экономической эффективности применения возобновляемых

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL dmitriy.karamov@mail.ru (A. 1); yumashevgeny@mail.ru (A. 2)

ORCID: 0000-0001-5360-4826 (A. 1); 0009-0003-8736-0155 (A. 2)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.17

источников энергии. Анализ упомянутых показателей можно выполнять различными способами.

2. Обзор математических моделей солнечной радиации

В данном разделе статьи представлены наиболее распространённые подходы к определению актинометрического потенциала на рассматриваемой территории.

Последовательность развития наиболее значимых зарубежных эмпирических моделей солнечной радиации имеет следующий вид: *P.I. Cooper* (1969), *R.E. Schulze* (1976), *M.A. Atwater* (1978), *M. Iqbal* (1980), *R.E. Bird* (1981), *G.L. Powel* (1984), *ASHRAE* (1985), *C. Gueymard* (1993) и многих других [1-10]. Представленные математические модели позволяют определять актинометрические показатели для условий ясного неба. В зарубежной литературе такие модели получили название *clear sky solar radiation model*.

Советская школа моделирования солнечной радиации начинается с работ: *С.И. Савинова* (1925), *В.Г. Кастрова* (1935), *В.Н. Украинцева* (1938), *Н.М. Копылова* (1949), *Т.Г. Берлянд* (1955), *С.И. Сивкова* (1968), *К.Я. Кондратьева* (1975) и других исследователей [11-17]. Дальнейшее совершенствование советской школы моделирования солнечной радиации получило развитие в работах современников. В настоящее время можно выделить следующих ученых активно развивающих данное направление, а именно: *А.А. Николаева*, *Б.В. Лукутина*, *С.Г. Обухова*, *В.А. Шакирова*, *А.Г. Васькова*, *В.А. Шкляева*, *А.А. Акулинина*, *И.А. Плотникова*, *Н.А. Халгаеву*, *Л.Г. Муста* и многих других [18-26].

Важным этапом при моделировании суммарной радиации является учет динамики облачности на рассматриваемой локации. Ранее отмечалось, что облачность может быть задана количеством солнечных и пасмурных дней по месяцам года либо реальными многолетними массивами фактической облачности. В частности, общая облачность в процентах, состав облаков, наличие облаков плохой погоды, а также высота их образования. Данные показатели выступают в качестве исходной информации для определения коэффициентов ослабления солнечной радиации. Наиболее известными эмпирическими моделями учета облачности являются: *В. Haurwitz* (1948), *S.B. Idso* (1969) *M.A. Atwater* (1974), *J.A. Davies* (1975), *J.E. Hay* (1977), *F. Kasten - G. Czeplak* (1980) и другие [27-36].

Следует отметить, что возможны различные сочетания представленных математических моделей солнечной радиации для условий ясного неба и учета фактической облачности. Эти сочетания в первую очередь зависят от объема данных метеорологических наблюдений, а также уровня их детализации [37].

Подробный литературный обзор, сопоставление и анализ эмпирических математических моделей солнечной радиации с учетом облачности представлены в работах: *В.А. Шакирова*, *L.T. Wong*, *A.E. Gürel* [38-40].

3. Методика расчета солнечной радиации

Поступающая на поверхность Земли солнечная радиация непостоянна и зависит от множества неизменяющихся и переменных параметров.

К неизменяющимся из года в год параметрам относятся: географические координаты и часовой пояс местности, количество дней расчетного периода, время восхода, зенита и заката Солнца, продолжительность дня и т. д.

К переменным параметрам в первую очередь необходимо отнести следующие показатели: наличие облачности, состав облаков, альbedo облаков и альbedo поверхности Земли, атмосферная масса с учетом изменяющегося давления и температуры воздуха, изменение толщины вертикального столба озонового слоя, изменение спектрального состава области видимого света, наличие в облаках водяного пара, аэрозолей и т. д.

4. Определение продолжительности дневного солнечного сияния в течение года

При расчете солнечной радиации в первую очередь для каждого дня определяется время восхода, зенита и заката Солнца на рассматриваемой территории.

Номер дня

$$n = n_1 - (n_2 \cdot n_3) + N_d - 30, \quad (1)$$

$$n_1 = (275 \cdot N_m / 9), \quad (2)$$

$$n_2 = (N_m + 9) / 12, \quad (2.5)$$

$$n_3 = 1 + \left((N_y - 4 \times (N_y / 4) + 2) / 3 \right), \quad (3)$$

где N_d – номер дня в году, N_m – номер месяца, N_y – номер расчетного года.

Средняя аномалия Солнца

$$M_s = (0,9856 \cdot t_i) - 3,289, \quad (4)$$

если восход

$$t_i = n + \left((6 - (lon/15)) / 24 \right), \quad (5)$$

если закат

$$t_i = n + \left((18 - (lon/15)) / 24 \right), \quad (6)$$

Истинная долгота Солнца

$$L_s = M_s + (1,916 \cdot \sin(M_s)) + (0,020 \cdot \sin(2 \cdot M_s)) + 282,634, \quad (7)$$

Прямое восхождение Солнца

$$R_s = \left(\tan^{-1}(0,91764 \times \tan(L_s)) + 360 \right) / 15, \quad (8)$$

Склонение Солнца

$$\sin dec = 0,39782 \cdot \sin(L_s), \quad (9)$$

$$\cos dec = \cos\left(\sin^{-1}(\sin dec)\right), \quad (10)$$

Местный часовой угол Солнца

$$\cos H_s = \left(\cos^{-1} \left(\frac{\sin dec \cdot \sin(lat)}{\cos dec \cdot \cos(lat)} \right) \right), \quad (11)$$

$$H_s = \left(360 - \cos^{-1}(\cos(H)) \right) / 15. \quad (12)$$

Местное время восхода и захода Солнца

$$T_s = H_s + R_s - (0,06571 \cdot t_i) - 6,622. \quad (13)$$

Перевод полученных значений на долготу местности

$$U_t = T_s - (lon / 15). \quad (14)$$

Определение времени восхода и захода Солнца на рассматриваемой территории

$$L_t = U_t + UTC \pm 24. \quad (15)$$

В качестве примера на рисунке 1 показаны временные характеристики восхода и захода Солнца в течение года для различных локаций, расположенных на территории Южного федерального округа, восточной Сибири и Дальнего Востока России.

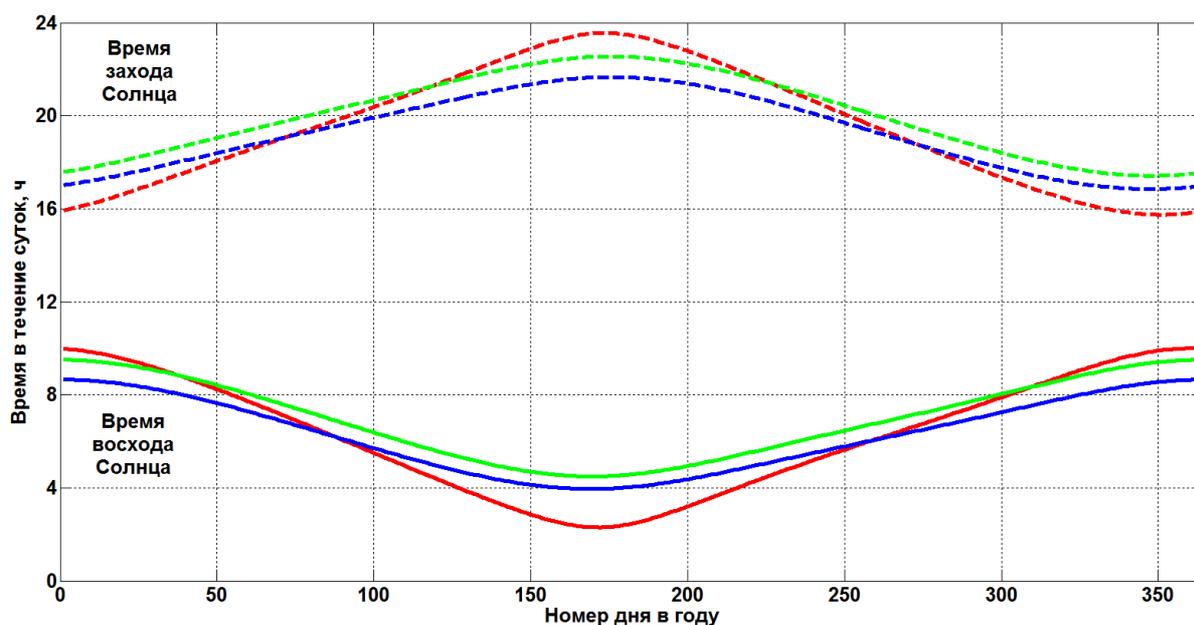


Рисунок 1: Время восхода и захода Солнца в течение года

Результаты, полученные при использовании WPS-сервиса, имеют характерный вид для рассматриваемого региона с максимальными значениями в весенне-летний периоды.

5. Математическое моделирование солнечной радиации с учетом динамики масс в нижних и средних слоях атмосферы

После того как определены временные интервалы, в рамках которых местный часовой угол Солнца больше нуля, а именно $H_s(t) > 0$, то производится расчет прямой, рассеянной и суммарной солнечной радиации на рассматриваемой территории. В работе используются основные положения, представленные в математической модели автора *M. Iqbal* 1980 года [41, 42].

$$\dot{I}_n(t) = 0,9751 \cdot E_0(t) \cdot \dot{I}_{sc}(t) \cdot \tau_r(t) \cdot \tau_o(t) \cdot \tau_g(t) \cdot \tau_w(t) \cdot \tau_a(t), \quad (16)$$

где E_0 – корреляционный коэффициент эксцентриситета орбиты Земли

$$E_0 = 1,00011 + 0,0034221 + 0,00128 \cdot \sin \Gamma + 0,000719 \cdot \cos 2\Gamma + 0,000077 \cdot \sin 2\Gamma, \quad (17)$$

$$\Gamma = 2\pi \cdot ((N_d - 1) / 365), \quad (18)$$

где $\tau_r(t), \tau_o(t), \tau_g(t), \tau_w(t), \tau_a(t)$ являются коэффициентами пропускания света, которые представлены Релеевским рассеиванием, рассеиванием в озоновом слое, наличием газа и капель воды в атмосфере, а также аэрозолями.

$$\tau_r(t) = e^{-0,0903 \times m_a^{0,84}(t) \times (1 + m_a(t) \cdot m_a^{1,01}(t))}, \quad (19)$$

$$\tau_o(t) = 1 - \left[\frac{0,1611 \cdot U_3(t) \cdot (1 + 139,48 \cdot U_3(t))^{-0,3035}}{-0,002715 \cdot U_3(t) \cdot (1 + 0,044 \cdot U_3(t) + 0,0003 \cdot U_3^2(t))^{-1}} \right], \quad (20)$$

$$\tau_g(t) = e^{-0,0127 \cdot m_a^{0,26}(t)}, \quad (21)$$

$$\tau_w(t) = 1 - 2,4959 \cdot U_1(t) \cdot \left[(1 + 79,034 \cdot U_1(t))^{0,6828} + 6,385 \cdot U_1(t) \right]^{-1}, \quad (22)$$

$$\tau_a(t) = e^{-l_{ao}^{0,873}(t) \cdot (1 + l_{ao}(t) \cdot l_{ao}^{0,7808}(t)) \cdot m_a^{0,9708}(t)}, \quad (23)$$

где $m_a(t)$ – атмосферная масса с учетом поправки на давление окружающей среды

$$m_a(t) = m_z(t) \cdot (p_{air}(t) / 1013,25) \quad (24)$$

где $p_{air}(t)$ (мбар) – местное атмосферное давление и атмосферная масса

$$m_z(t) = 1 / \sin \delta(t). \quad (25)$$

Угол склонения Солнца

$$\delta(t) = \sin^{-1} \left(\sin(23,45) \cdot \sin(360 / 365 \cdot (n_d(t) - 81)) \right). \quad (2.29)$$

Угол возвышения Солнца

$$\alpha(t) = 90 - lat + \delta(t). \quad (26)$$

Оптическая длина пути в озоновом слое атмосферы с учетом поправки на давление окружающей среды (единицы Добсона, (еД))

$$U_3(t) = l_{oz}(t) \cdot m_a(t), \quad (27)$$

где $l_{oz}(t)$ – (еД) вертикальная величина озонового слоя.

Относительная длина оптического пути с учетом наличия капель воды и водяного пара в атмосфере, см

$$U_1(t) = w(t) \cdot m_z(t), \quad (28)$$

где w (см) – осажденная толщина водяного пара и капель воды в атмосфере. $w(t)$ рассчитывается с учетом поправки на температуру и давление окружающей среды w' (см)

$$w(t) = w'(t) \cdot (p_{air}(t) / 1013,25)^{3/4} \cdot (273 / T_{air}(t))^{1/2}. \quad (29)$$

Оптическая глубина аэрозолей

$$I_{ao}(t) = 0,2758 \cdot I_{ao;\lambda|\lambda = 0,38\mu m}(t) + 0,35 \cdot I_{ao;\lambda|\lambda = 0,5\mu m}(t). \quad (30)$$

AErosol RObotic NETwork (AERONET) позволяет получить многолетние данные по $w'(t)$, $I_{ao;\lambda|\lambda = 0,38\mu m}(t)$, $I_{ao;\lambda|\lambda = 0,5\mu m}(t)$. Ежегодно метеостанции, обладающие соответствующим оборудованием для измерения данных параметров, передают накопленную информацию в международный центр, где полученные данные хранятся и систематизируются.

Прямая солнечная радиация

$$\dot{I}_b(t) = \sin \alpha(t) \cdot \dot{I}_n(t) \cdot C_{c-dr}(t), \quad (31)$$

где $\sin \alpha$ (градусы) угол возвышения и C_{c-dr} коэффициент уменьшения прямой солнечной радиации за счет облачности. При определении коэффициента ослабления солнечной радиации используется математическая модель авторов *Fritz Kasten* и *Gerhard Czeplak* модификации 1980 года.

$$C_{c-dr}(t) = 1 - 0,75 \cdot (N(t)/8)^{3,4}, \quad (32)$$

где N зависит от степени покрытия неба облаками в определенный момент времени (3 часовой интервал) – измеряется в октантах (oktas) [32].

Рассеянная солнечная радиация образуется в результате рассеивания солнечных лучей в нижних слоях атмосферы и отражённая от земной поверхности земли.

$$\dot{I}_d(t) = (\dot{D}_r(t) + \dot{D}_a(t) + \dot{D}_m(t)), \quad (33)$$

где $\dot{D}_r(t)$ – Релеевское рассеивание солнечной радиации после первого прохождения атмосферы.

$$\dot{D}_r(t) = \frac{0,79 \cdot \dot{I}_{sc}(t) \cdot \sin \alpha(t) \cdot \tau_o(t) \cdot \tau_g(t) \cdot \tau_w(t) \cdot \tau_{aa}(t) \cdot 0,5 \cdot (1 - \tau_r(t))}{1 - m_a(t) + m_a^{1,02}(t)}, \quad (34)$$

$$\tau_{aa}(t) = 1 - (1 - \omega_0) \cdot (1 - m_a(t) + m_a^{1,06}(t)) \cdot (1 - \tau_a(t)). \quad (35)$$

где ω_0 – безразмерная величина, характеризующая альбедо при однократном рассеянии приходящей энергии Солнца через толщу аэрозолей в средних слоях атмосферы, принимается равной 0,9 [[39]].

Аэрозольное рассеивание солнечной радиации при проходе через атмосферу

$$\dot{D}_a(t) = \frac{0,79 \cdot \dot{I}_{sc}(t) \cdot \sin \alpha(t) \cdot \tau_o(t) \cdot \tau_g(t) \cdot \tau_w(t) \cdot \tau_{aa}(t) \cdot 0,84 \cdot (1 - \tau_r(t))}{1 - m_a(t) + m_a^{1,02}(t)} \quad (36)$$

$$\tau_{as}(t) = \tau_a(t) / \tau_{aa}(t). \quad (37)$$

Рассеянная солнечная радиация, многократно отраженная между поверхностью Земли и нижними слоями атмосферы,

$$\dot{D}_m(t) = \frac{(\dot{I}_n(t) \cdot \sin \alpha(t) + \dot{D}_r(t) + \dot{D}_a(t)) \cdot \rho_g(t) \cdot \rho_a(t)}{1 - \rho_g(t) \cdot \rho_a(t)}, \quad (38)$$

Альбедо нижних слоев атмосферы с учетом динамики изменения облачности

$$\rho_a(t) = 0,0685 + (1 - \rho_a^{ct}(t)) \cdot (1 - \tau_{as}(t)), \quad (39)$$

где $\rho_g(t)$ - альбедо земной поверхности, о.е. $\rho_a^{ct}(t)$ – альбедо облаков относительно каждого временного интервала, о.е. Альбедо облаков зависит от их типа и состава. В данной работе

используется следующий алгоритм для определения данного параметра. Для каждого временного интервала формируется массив данных в котором представлен актуальный состав облаков, а также их процентное соотношение.

Суммарная солнечная радиация, Вт/м²

$$\dot{I}_t(t) = \dot{I}_b(t) + \dot{I}_d(t). \quad (40)$$

Таким образом предложенный тандем моделей для расчета суммарной солнечной радиации в зависимости от географического положения местности, фактических значений природно-климатических показателей, динамики облачности, водяного пара и аэрозолей в нижних и средних слоях атмосферы является универсальным и может быть использован для любой локации. В качестве примера на рисунке 2 показана характеристика суммарной солнечной радиации, рассчитанная по вышеописанной методике для населенного пункта «Иннялы» Республика Саха (Якутия) по данным последних 14 лет метеорологических наблюдений.

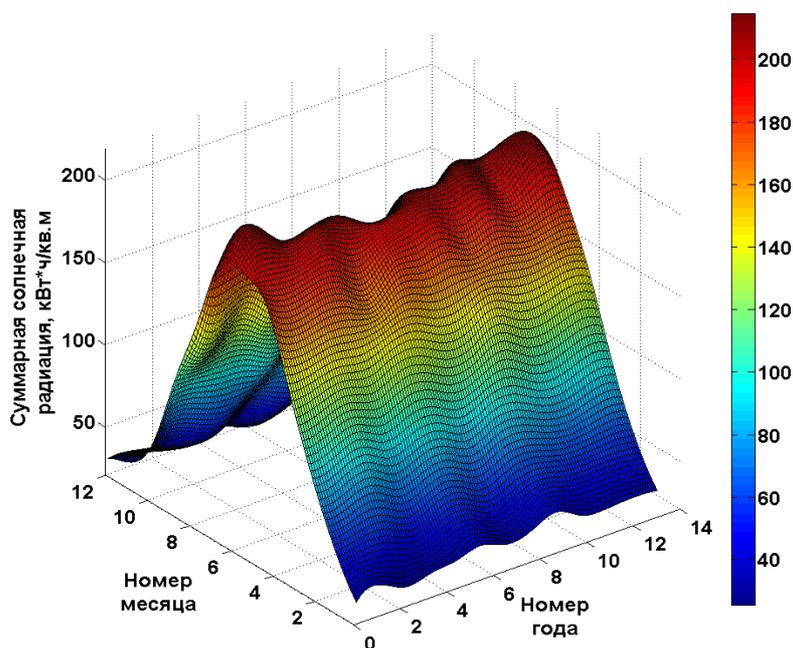


Рисунок 1: Суммарная солнечная радиация для н.п. «Иннялы»

На основании полученной характеристики можно определить следующие годовые показатели: минимальное значение – 1240 кВт·ч/м²; максимальное значение – 1325 кВт·ч/м²; среднее значение – 1285 кВт·ч/м². Если сравнить полученные результаты моделирования с реальными значениями с рассматриваемой локацией, то среднегодовая солнечная радиация составляет 1278 кВт·ч/м².

6. Заключение

В статье представлена универсальная методика расчета солнечной радиации на рассматриваемой территории. Методика включает в себя основные метеорологические параметры необходимые для расчета прямой, рассеянной, суммарной солнечных радиаций.

В настоящий момент авторы статьи совместно с коллегами из ИДСТУ СО РАН разрабатывают WPS-сервис, позволяющий реализовать данную методику с возможностью формирования типичного, оптимистичного, пессимистичного метеорологического года. Более того усилия будут направлены на разработку прогнозной модели на основании ретроспективных данных природно-климатических показателей, а также ряда астрономических параметров.

7. Благодарности

Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды», рег. № 124052100088-3).

8. Список источников

- [1] Cooper P.I. The absorption of radiation in solar stills. *Solar energy*, 1969, vol. 12, pp. 333–346.
- [2] Schulze R.E. A physically based method of estimating solar radiation from suncards. *Agricultural Meteorology*, 1976, vol. 16, pp. 85–101.
- [3] Atwater M.A., Ball J.T. A surface solar radiation model for cloudy atmosphere. *Monthly weather review*, 1980, vol. 109, pp. 878–888.
- [4] Iqbal M. Prediction of hourly diffuse solar radiation from measured hourly global radiation on a horizontal surface. *Solar energy*, 1980, vol. 24, pp. 491–503.
- [5] Bird R.E., Hulstrom R.L. A simplified clear sky model for direct and diffuse insolation on horizontal surfaces. *SERI Technical Report*, 1981, vol. 69, pp. 642–761.
- [6] Bird R.E. A simple, solar spectral model for direct-normal and diffuse horizontal irradiance. *Solar energy*, 1984, vol. 32, pp. 461–471.
- [7] Powel G.L. Comparative evaluation of SOLMET clear sky global solar radiation model. *Solar energy*, 1984, vol. 32, pp. 453–460.
- [8] Carroll J. J. Global transmissivity and diffuse fraction of solar radiation for clear and cloudy skies as measured and as predicted by bulk transmissivity models. *Solar energy*, 1985, vol. 35, pp. 115–118.
- [9] Barbaro S., Cannistraro G., Giaconia C., Orioli A. The ASHRAE clear sky model. An evaluation in the Mediterranean Zone. *Solar & Wind Technology*, 1984, vol. 5, pp. 111–116.
- [10] Gueymard C. Critical analysis and performance assessment of clear-sky solar-irradiance models using theoretical and measured data. *Solar energy*, 1993, vol. 51, pp. 121–138.
- [11] Савинов С.И. Солнечная земная и атмосферная радиация // Климат и погода. – 1926. – № 2-3. – С. 12-59.
- [12] Кастров В. Г. К вопросу об основной актинометрической формуле. // Метеорол. Вестник. – 1928. – № 7. – С. 173-175.
- [13] Украинцев В. Н. Приближенное вычисление сумм прямой и рассеянной солнечной радиации // Метеорология и гидрология. – 1939. – № 6. – С. 3-18.
- [14] Копылов Н. М. О приближенных вычислениях сумм солнечной радиации // Тр. ГГО. – 1949. – № 14(76). – С. 63-69.
- [15] Берлянд Т. Г. Методика климатологических расчетов суммарной радиации // Метеорология и гидрология. – 1960. – № 6. – С. 9-12.
- [16] Кондратьев К.Я. Актинометрия. Л.: Гидрометеиздат, 1965, 690 с.
- [17] Кондратьев К.Я., Пивоварова З.И., Федорова М.П. Радиационный режим наклонных поверхностей. Л.: Гидрометиздат, 1978. 170 с.
- [18] Николаев А.А. Климатические ресурсы солнечной радиации и ветра на территории Среднего Поволжья и возможности их использования в энергетике: дис. канд. геогр. наук. – 2000. – 139 с.
- [19] Николаев А.А. Косвенные методы расчета характеристик солнечной радиации // Вестник Удмуртского университета. – 2013. – № 1. – С 130-135.
- [20] Обухов С.Г., Плотников И.А. Математическая модель прихода солнечной радиации на произвольно-ориентированную поверхность для любого региона России. Международный научный журнал Альтернативная энергетика и экология. – 2017. - № 16-18. – С 43-56.

- [21] Шакиров В.А., Артемьев А.Ю. Учет данных метеостанций при анализе эффективности применения солнечных энергетических установок. Вестник Иркутского государственного технического университета. – 2015. - № 3. – С 227-232.
- [22] Васьков А.Г., Нарынбаев А.Ф. Обзор и классификация методов оценки и прогнозирования солнечной радиации // Вестник МЭИ. 2020. № 4. С. 49-61.
- [23] Исаков С.В., Шкляев В.А. Оценка поступления солнечной радиации на естественные поверхности с применением геоинформационных систем // Географический вестник. 2012. № 1. С. 72-80.
- [24] Акулинин А., Смыков В. Оценка возможностей солнечной энергетики на основе точных наземных измерений солнечной радиации // Проблемы региональной энергетики. 2008. № 1. С. 29-39.
- [25] Кошелев А.А., Халгаева Н.А. Оценка целесообразности и выбор оптимальных решений для солнечного энергоснабжения в условиях азиатской части России // Известия Российской академии наук. Энергетика. 2006. № 3. С. 102-109.
- [26] Муста Л.Г. Моделирование потока солнечной радиации на произвольно ориентированную поверхность // Вестник гражданских инженеров. 2009. № 1. С. 110-113.
- [27] Haurwitz, G.: Insolation in Relation to Cloud Type. *J. Met.*5, 110–113 (1948).
- [28] Idso S. B. Atmospheric attenuation of solar radiation // *Journal of Atmospheric Sciences*. – 1969. – Т. 26. – №. 5. – С. 1088-1095.
- [29] Atwater M. A., Brown Jr P. S. Numerical computations of the latitudinal variation of solar radiation for an atmosphere of varying opacity // *Journal of Applied Meteorology and Climatology*. 1974, vol. 13, pp. 289–297.
- [30] Davies J.A., Schertzer W., Nunez M. Estimating Global Solar Radiation. *Boundary-Layer Meteorology*. 1975 vol. 9, pp. 33–52.
- [31] Suckling P.W., Hay J.E. A cloud layer-sunshine model for estimating direct, diffuse and total solar radiation. *Atmosphere*. 1977, vol. 15, pp. 194–207.
- [32] Kasten F., Czeplak G. Solar and terrestrial radiation dependent on the amount and type of cloud. *Solar Energy*, 1980, vol. 75, pp. 177–189.
- [33] Ehnberg J.S.G., Bollen M.H.J. Simulation of global solar radiation based on cloud observations. *Solar Energy*, 2005, vol. 24, pp. 157–162.
- [34] Hamdan M.A. Gazzawi N. The effect of clouds on solar radiation. *Solar Energy*, 1993, vol. 34, pp. 29–32.
- [35] Dissing D., Wendler G. Solar Radiation Climatology of Alaska. *Theoretical applied climatology*, 1998, vol. 61, pp. 161–175.
- [36] Badescu V., Dumitrescu A. New models to compute solar global hourly irradiation from point cloudiness. *Energy Conversion and Management*, 2013, vol. 67, pp. 75–91.
- [37] Карамов Д.Н. Математическое моделирование солнечной радиации с использованием многолетних метеорологических рядов находящихся в открытом доступе // Известия томского политехнического университета. Инжиниринг георесурсов. 2017. Т. 328. № 6. С. 28–38.
- [38] Vladislav Shakirov, Irina Ivanova, Tatyana Tuguzova (2022) Development of empirical solar radiation models with genetic algorithm and extended validation procedure, *International Journal of Green Energy*,
- [39] Wong L.T., Chow W.K. Solar radiation model. *Applied Energy*, 2001, vol. 69, pp. 191–224.
- [40] Gürel A.E., Ağbulut Ü., Bakır H., Ergün A., Yıldız G. A state of art review on estimation of solar radiation with various models. *Heliyon*, 2023, vol. 9, e13167.
- [41] Iqbal M. Estimation of the monthly average of the diffuse component of total insolation on a horizontal surface. *Solar Energy* 1978; 20 (1): 101–5.
- [42] Iqbal M. An introduction to solar radiation. Toronto: Academic press, 1983.

Подход к развертыванию программного обеспечения вычислительного кластера по облачной модели

Роман Костромин

Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

Аннотация

В работе обсуждается модернизация модели управления вычислительным кластером с использованием контейнеризации для расширения возможностей суперкомпьютерного центра. Предлагаемый подход позволяет сохранить традиционные способы управления ресурсами кластера и внедрить новые сценарии использования. Ожидается, что сочетание методов и средств контейнеризации позволит существенно ускорить подготовку и проведение научных исследований, снизить затраты на содержание инфраструктуры и улучшить общую полезную загрузку кластера. Предлагаемая модель разрабатывается с расчетом на привлечение новых исследователей и разработчиков за счет расширения возможностей использования вычислительных ресурсов и внедрения передовых технологий.

Ключевые слова

Вычислительный кластер, администрирование, автоматизация развертывания операционных систем, контейнеризация

1. Введение

В последние годы наблюдается значительная тенденция к расширению сферы применения суперкомпьютеров, включая решение новых задач из области машинного обучения, искусственного интеллекта, микросервисного подхода и других передовых технологий [1, 2]. Это расширение подчеркивает не только сохраняющуюся актуальность суперкомпьютеров, но и их растущее значение в качестве мощного инструмента для развития современных научных и технологических направлений. Благодаря этому академические суперкомпьютерные центры (СКЦ) становятся ключевым элементом в реализации новых проектов и исследований, подтверждая свою роль как важного актива в достижении прогресса в многочисленных областях науки и техники [3].

Традиционно, СКЦ представляют собой комплексные системы, поставляемые специализированными подрядчиками и требующие значительных усилий для поддержания их работоспособности. В состав таких систем входят разнообразные инструменты и программное обеспечение (ПО), включая средства для параллельного выполнения команд, мониторинга, оповещения, а также для развертывания базовых образов операционных систем (ОС). Однако, существующие средства автоматизации часто ограничиваются поддержкой исходного состояния узлов, рекомендованного поставщиками, что влечет за собой ограничения на спектр выполняемых научных приложений и не предоставляет инструментов для пользователей, чьи задачи требуют специализированного окружения и определенных версий ПО.

С точки зрения решения упомянутых выше проблем, перспективным направлением является применение методов и средств управления ресурсами основанных на виртуализации и контейнеризации, применяемых в центрах обработки данных и у облачных провайдеров [1, 2].

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: roman@kostromin.net (A. 1)

ORCID: 0000-0001-8406-8106 (A. 1)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.18

К сожалению, особенности СКЦ, ориентированных на выполнение параллельных MPI приложений, накладывают ограничения на применение подходов, традиционных для облачных ресурсов и обоснованно требуют проведения дополнительных исследований в этой области.

Таким образом, в данной работе обсуждаются вопросы модернизации классического подхода управления вычислительным кластером (ВК) с целью интеграции методов и средств из сферы управления облачными ресурсами. Предложены основные принципы обновленного подхода, а также ключевые его компоненты. Прототип данного подхода был протестирован на резервных узлах Центра коллективного пользования «Иркутский суперкомпьютерный центр СО РАН» [4] и продемонстрировал реализуемость данного проекта и соответствие предъявляемым администраторами требованиям.

2. Особенности управления ВК

Классический цикл работы с ВК со стороны пользователя включает следующие этапы [5]:

- Удаленный вход на кластер;
- Копирование данных между ВК и компьютером пользователя;
- Редактирование исходных текстов программ;
- Компиляция программ;
- Запуск задач и работа с очередью;
- Копирование результатов вычислений;
- Завершение сеанса.



Рисунок 1: Типовая архитектура ВК с точки зрения пользователя

Иными словами, пользователь формирует паспорт задания с учетом регламента и квот доступа к кластеру, ставит его в очередь локального менеджера ресурсов (LPM) через узел доступа к кластеру (см. Рисунок 1). В соответствии с политикой диспетчирования задания распределяются по вычислительным узлам. Узлы взаимодействуют между собой в процессе вычислений по высокоскоростной сети, такой как Infiniband или аналогичной. С этой точки зрения ВК воспринимается как черный ящик, способы взаимодействия с которым строго регламентированы. Например, установка дополнительного ПО допустима только в домашний каталог пользователя силами самих пользователей.

Иной взгляд на ВК имеют его администраторы (см. Рисунок 2). Такая архитектура ВК встречается повсеместно в СКЦ, поставляемых российскими суперкомпьютерными компаниями. В ней выделяется ключевая роль виртуальной машины (VM) master, для которой настроено реплицирование файловой системы виртуального диска через технологию DRBD на два сервера (master-1 и master-2). VM master управляет сетью ВК, вычислительными узлами, содержит локальный менеджер ресурсов (LPM), каталоги пользователей LDAP, систему мониторинга и т.д. Доступ к кластеру осуществляется через основной и резервный узлы доступа (access-1 и access-2). Все системы подключены по высокоскоростной сети к производительной системе хранения данных (СХД).

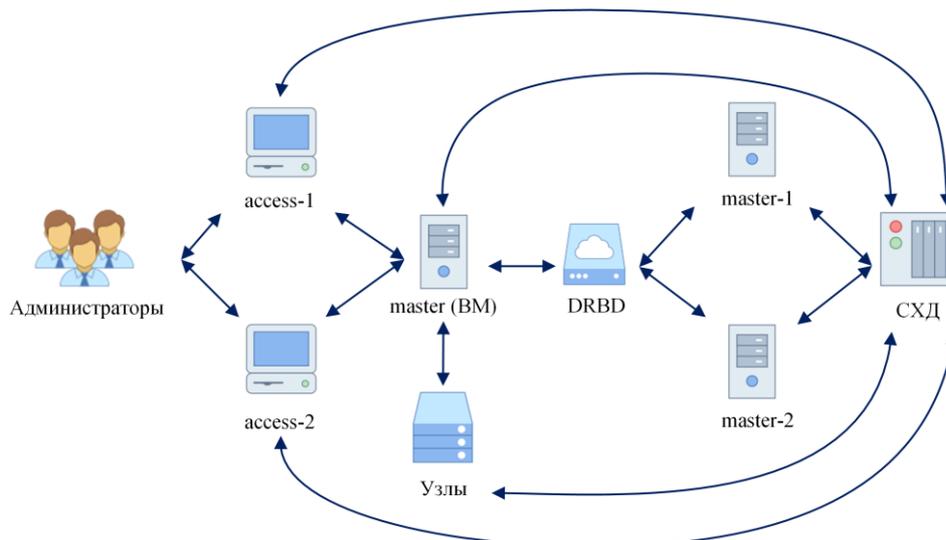


Рисунок 2: Типовая архитектура ВК с точки зрения администратора

Рассматриваемая архитектура хорошо зарекомендовала себя в течение длительного времени эксплуатации ВК, но при этом выявлены следующие недостатки, а именно:

- повреждение структуры DRBD или виртуального диска master приводит к выходу из строя ВМ master;
- при выходе из строя ВМ master затрудняется управление кластером и его использование из-за проблем синхронизации SSH и LDAP;
- при выходе из строя СХД данных затрудняется управление кластером и его использование (нет каталогов пользователей, библиотек и программ);
- узлы master-1, master-2, access-1 и access-2 большую часть времени простаивают.

Как показала практика, конфигурация с DRBD может длительное время работать надежно, но в случае сбоев в работе ВМ master восстановление и устранение неполадок занимает значительное время, что ведет к простоям работы ВК. Кроме того, регулярные запросы пользователей о возможности установки дополнительных системных и прикладных библиотек, отсутствующих на кластере, актуализируют задачу модернизации существующей модели работы ВК. Таким образом, несмотря на достоинства архитектуры классического ВК, необходимо принять меры по устранению выявленных недостатков, чтобы обеспечить более стабильное и эффективное функционирование кластера в будущем, а также создать условия для привлечения новых пользователей.

Для решения этих проблем облачные провайдеры предлагают два подхода – применение виртуализации и контейнеризации. Преимущества виртуализации управляющей системы заключаются в том, что благодаря разделению ролей ВМ master на несколько независимых ВМ, снижаются риски одновременного отказа всех компонентов управляющей системы. При этом отказоустойчивость обеспечивается созданием кластера виртуализации с применением нескольких физических серверов с конфигурированием живой миграции и реплицирования ВМ между узлами. Применение виртуализации целесообразно для разделения управляющих компонентов на несколько независимых виртуальных сущностей, однако для вычислительных узлов виртуализация влечет значительные накладные расходы [6].

Как показано в работах [7-8], популярным и устоявшимся подходом к построению изолированных вычислительных сред является контейнеризация ПО. В частности, такая контейнеризация осуществляется с помощью Docker. В отличие от виртуализации, контейнеризация снижает затраты на подготовку, хранение и развертывание настроенных образов ОС и ПО [8]. Однако Docker это не лучшее решение для высокопроизводительных сред, в отличие от Singularity, которая обеспечивает полную поддержку образов Docker и устраняет недостатки Docker для запуска контейнеров с точки зрения производительности и обеспечения безопасности вычислений [9].

Для управления группой узлов в среде контейнеризации (оркестрации) традиционно применяют Kubernetes, на его основе и предложена новая модель управления СКЦ (см. Рисунок 3). Ее основой является управляющий кластер на базе Kubernetes, работающий на нескольких узлах управления (УУ). Средствами Kubernetes выполняется запуск образов контейнеров (Docker и Singularity), хранимых в локальной СХД. Часть контейнеров содержат компоненты управления (КУ) кластером, которые выполняются на УУ. Kubernetes создает изолированные сети (виртуальная сеть) для организации взаимодействия между контейнерами вычислительных задач (КВЗ). На вычислительных узлах осуществляется запуск контейнеров Singularity с поддержкой локальных менеджеров ресурсов (PBS, Slurm) в рамках установленных квот пользователей для работы с очередью заданий. Образы, хранимые в СХД подготавливаются администраторами СКЦ заранее и обеспечивают послойную интеграцию необходимых компонентов для приложений и их модулей, запускаемых на ВК.

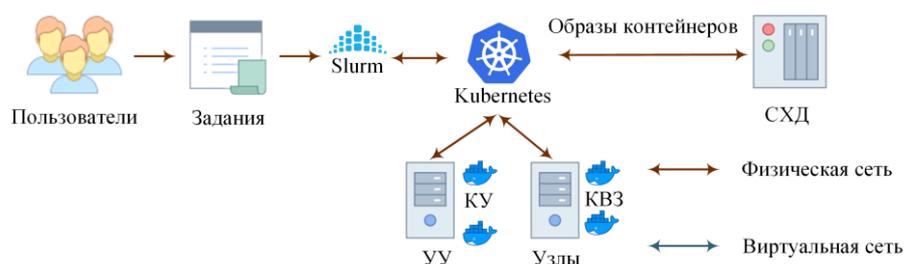


Рисунок 3: Контейнеризация компонентов ВК

3. Вопросы совместимости предложенного подхода с классическим подходом управления ВК

При обновлении модели управления кластером необходимо помнить, что СКЦ ориентирован в первую очередь на пользователей и применение кластера для научных задач не должно быть связано со значительными техническими сложностями. Поэтому необходимо обеспечить как преемственность (поддержку выполнения приложений, несовместимых с новыми системными библиотеками), так и поддержку новых библиотек и технологий. В рамках одной ОС реализовать оба фактора затруднительно. Одним из решений является развертывание (provisioning или deploy) нужной версии ОС и всех необходимых компонентов на узлах, подходящих для задач пользователей. Этот процесс хорошо автоматизирован с помощью специализированных средств (xCAT, OpenHPC, MAAS). Но с точки зрения полезной загрузки ресурсов данный подход несостоятелен, т.к. каждое развертывание занимает порядка 20 минут, что совершенно не применимо для ВК.

По этой причине стоит рассматривать узлы в режиме stateless, т.е. ОС узла не имеет определенного состояния и настроек, загружается по сети и не требует для своей работы физического накопителя, является легковесной и занимает минимальный объем оперативной памяти (до 500 МБ). Развертывание такой системы занимает порядка 5 минут, из которых до 80% затрачивается на этап загрузки BIOS и самотестирование узла до начала непосредственной загрузки ОС. Основное назначение такой системы – получить от управляющего узла образ контейнера из СХД по высокоскоростной сети, запустить его (10-15 сек.) и передать все доступные ресурсы узла данному контейнеру. Именно Kubernetes реализует управление контейнерами таких бездисковых ОС. Контейнер после завершения работы уничтожается и узел полностью освобождается. В рамках предложенного подхода Kubernetes встраивается в существующий менеджер ресурсов (PBS, Slurm), благодаря чему сохраняется подход с общей очередью заданий и учет квот пользователей для тех, кому в явном виде применение контейнеризации не требуется.

Применение контейнеров позволяет создавать окружение, которое требуется для заданий пользователей. Администраторы СКЦ подготавливают базовые образы систем, содержащие все

необходимые компоненты для разных типов задач (C++ и MPI, Python и CUDA, Hadoop, Apache Ignite и т.д.). Образцы контейнеров могут содержать как устаревшие версии компиляторов и системных библиотек, так и самые современные. Благодаря этому расширяется спектр поддерживаемых языков и системных библиотек, сохраняется привычный способ работы с кластером, снижаются накладные расходы на развертывание ОС на узлах и сокращается количество потенциальных точек отказа системной составляющей кластера.

Для тестирования приложенного подхода было задействовано 10 узлов ВК со следующими характеристиками (см. Таблица 1).

Таблица 1. Характеристики узлов

CPU	RAM	Storage	Network	OS
Intel Xeon E5-2695 v4 «Broadwell» 2.1 GHz, (2 CPU, 72 cores)	128 Gb	256 Gb SSD; 40 Tb, RAID 5	Ethernet, 1 Gbit/s; InfiniBand, 40 Gbit/s	Linux-based

Средствами xCAT на данных узлах была развернута RedHat-совместимая ОС RockyLinux 9.3. Она является продолжателем ОС Centos, чья поддержка завершена в 2024 г. Для запуска MPI-приложений подготовлен базовый образ (БО) контейнера, в который интегрированы требуемые системные модули (СМ) и компоненты LMP. На основе БО созданы дополнительные образы для разных классов ресурсов и типов задач. В них дополнительно встроены средства работы с временной базой данных (ВБД), средства интеграции с системой мониторинга и локальным Git-репозиторием.



Рисунок 4: Структура слоев контейнеров для ВК

Дополнительно настроена VM с Kubernetes и выполнена настройка тестовых узлов для взаимодействия с Kubernetes. Через LMP Slurm поставлено в очередь задание, на основании которого средствами Kubernetes выполнено развертывание образов с последующим запуском MPI-задач для оценки производительности узлов (Linpack). Аналогичный набор задач был запущен на данных узлах в явном виде без применения средств контейнеризации. В обоих случаях производительность составила 7.9 Tflops. Тесты производительности Singularity на большем количестве узлов подтверждают, что применение Singularity для ВК не оказывает негативного влияния на общую производительность Tflops [10, 11].

4. Заключение

Применение контейнеризации является востребованным направлением для интеграции в СКЦ. Однако необходимо этот подход к контейнеризации развивать комплексно, обеспечив поддержку личного кабинета пользователя кластера и возможность самостоятельной тонкой настройки базовых образов и т.д., т.к. существующие решения не покрывают всего спектра

применения контейнеризации для ВК. При этом сам подход является гибким и тестирование на небольшом количестве узлов подтверждает его функциональность. Данное исследование находится на начальном этапе, но первые результаты показывают перспективность и достижимость поставленных целей, формируя достаточную основу для дальнейших исследований.

5. Благодарности

Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации, проект № FWEW-2021-0005 «Технологии разработки и анализа предметно-ориентированных интеллектуальных систем группового управления в недетерминированных распределенных средах» (рег. № 121032400051-9).

6. Список литературы

- [1] А.О. Кудрявцев, В.К. Кошелев, А.О. Избышев, А.И. Аветисян. Высокопроизводительные вычисления как облачный сервис: ключевые проблемы, Параллельные вычислительные технологии (ПаВТ 2013): Труды междунар. науч. конф., Челябинск, Издательский центр ЮУрГУ, 2013, с. 432–438.
- [2] S. Deng. Cloud-Native Computing: A Survey From the Perspective of Services, Proceedings of the IEEE, 2023. vol. 112, pp. 12–46. DOI: 10.1109/JPROC.2024.3353855.
- [3] С.В. Поляков, А.В. Выродов, Д.В. Пузырьков, М.В. Якововский. Облачный сервис для решения многомасштабных задач нанотехнологии на суперкомпьютерных системах, Труды ИСП РАН, 2015, Т. 27, № 6, с. 409–420. DOI 10.18522/2311-3103-2016-12-103114.
- [4] Иркутский суперкомпьютерный центр СО РАН [Электронный ресурс]. Режим доступа: <https://hpc.isc.ru>. (Дата обращения 10.06.2024 г.).
- [5] С.А. Жуматий, К.С. Стефанов. Суперкомпьютеры: Администрирование, М: МАКС Пресс, 2018.
- [6] В.М. Shabanov, O.I. Samovarov. Building the Software-Defined Data Center, Programming and Computer Software, 2019, vol. 45, pp. 458–466. DOI: 10.1134/S0361768819080048.
- [7] А.О. Кудрявцев и др. Разработка и реализация облачной системы для решения высокопроизводительных задач, Труды ИСП РАН, 2013, Т. 24, с. 13–34.
- [8] М.А. Rodriguez, B. Rajkumar. Container-based cluster orchestration systems: A taxonomy and future directions, Software: Practice and Experience, 2018, vol. 49, pp. 698–719. DOI: 10.1002/spe.2660.
- [9] O. Freyermuth, P. Wienemann, P. Bechtle, K. Desch. Operating an HPC/HTC cluster with fully containerized jobs using HTCondor, Singularity, CephFS and CVMFS, Computing and Software for Big Science, 2021, vol. 5, is. 1, pp. 9. DOI: 10.1007/s41781-020-00050-y.
- [10] G. Hu, Y. Zhang, W. Chen. Exploring the performance of singularity for high performance computing scenarios, 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPC/SmartCity/DSS), IEEE, 2019, pp. 2587–2593.
- [11] А.Г. Феоктистов, Р.О. Костромин, М.Л. Воскобойников, Д.И. Ли-Дэ. Организация вычислительной среды разработки и применения научных рабочих процессов на основе контейнеризации, Вычислительные технологии, 2023, Т. 28, № 6, с. 151–164.

Оптимальное планирование долгосрочных режимов работы каскадов ГЭС в целях максимизации дохода генерирующей компании

Алексей Михеев¹, Евгений Осипчук¹, Николай Абасов¹, Вячеслав Никитин¹

¹ Институт систем энергетики им. Л.А. Мелентьева СО РАН, ул. Лермонтова, 130, Иркутск, 664033, Россия

Аннотация

В статье представлен подход к согласованному планированию долгосрочных режимов работы гидроэлектростанций (ГЭС), принадлежащих целевой генерирующей компании и работающих в нескольких каскадах. Планирование оптимальных режимов осуществляется в целях максимизации ожидаемого совокупного дохода компании за расчетный период на основании среднемесячного прогноза приточности водохранилищ и динамики средних цен на оптовом рынке электрической энергии с учетом гидрологических, водохозяйственных, экологических и энергетических ограничений, а также графика плановых ремонтов основного генерирующего оборудования ГЭС. Для нахождения максимума нелинейной целевой функции при заданных ограничениях предложена и реализована комбинация методов линейного и динамического программирования, когда сначала выполняется поиск допустимого, но достаточно близкого к оптимальному решения, а потом производится его уточнение с приемлемой точностью. Предварительные оптимизационные расчеты показывают, что предлагаемый метод позволяет рационально определить оптимальные параметры управления – общие среднемесячные расходы через створы ГЭС в каскадах с учетом динамики зависимостей рыночных цен на электроэнергию от возможных совокупных объемов ее генерации в энергосистеме.

Ключевые слова

гидроэнергетика, каскад ГЭС, долгосрочное планирование, нелинейная оптимизация, двухэтапная оптимизация, динамика цен электроэнергии

1. Введение

Развитие рынка электрической энергии, как одного из действенных механизмов регулирования спроса и предложения в электроэнергетике, ставит новые вызовы и задачи перед его основными участниками, в том числе и перед генерирующими компаниями негосударственного сектора, для которых большое значение имеет финансово-экономическая эффективность. Это особенно важно для энергосистем со значительной долей ГЭС с водохранилищами многолетнего регулирования, например, таких как Объединенная энергетическая система (ОЭС) Сибири, которые позволяют делать значительные запасы электроэнергии посредством аккумулирования водных ресурсов и, соответственно, имеют возможности для более гибкого управления ими. В связи с этим перед компаниями неизбежно встает задача обеспечения максимальной доходности от конкурентных поставок электроэнергии на оптовый рынок, но с обязательным выполнением всех многочисленных гидрологических, водохозяйственных, экологических и энергетических ограничений на работу ГЭС [1].

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: avmiheev@gmail.com (A. 1); eugene.os@mail.ru (A. 2); abasov_nv@rambler.ru (A. 3); nikitin1310@mail.ru (A.4)

ORCID: 0000-0003-4219-8812 (A. 1)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.19

Проблема эффективного долгосрочного управления каскадами ГЭС остается актуальной в мировой повестке и представляет предмет интереса в странах с развитой гидрогенерацией - таких как Китай, США, Бразилия и Россия, во-первых, в связи с ее комплексностью влияющих на гидроэнергетику природообусловленных и социально-экономических факторов, во-вторых, в связи с продолжающимися вводами новых больших ГЭС и формированием новых гидроэнергетических каскадов, имеющих значительное влияние на работу национальных энергосистем. Большинство работ рассматривают преимущественно оптимизацию технико-экономической эффективности каскадов ГЭС, которая выражается в максимизации произведенной электроэнергии при прогнозируемых на определенном горизонте времени водных запасах [2, 3].

Для решения подобных задач часто используется метод динамического программирования [4-6] в том числе его модификации [7]. Применяется аппарат математического (линейного и нелинейного) программирования [8-9]. В последнее время предприняты попытки использовать аппарат нейронных сетей с глубоким обучением [10-11]. Авторами [12] разработан нелинейный метод долгосрочного оптимального управления каскадными гидроэлектростанциями с заданными типами кривых цен на электроэнергию. Нелинейность используемой расчетной модели каскада ГЭС авторы сводят методом полиномиальной аппроксимации к разрешимой модели смешанного целочисленного линейного программирования. Однако, максимизация дохода отдельной генерирующей компании не рассматривается.

2. Математическое моделирование долгосрочных режимов работы каскадов ГЭС с учетом прогноза приточности, цен на оптовом рынке электроэнергии

В данной работе используется гидрологическая модель, описанная в работе [13], где изменение объема или скорость сработки/наполнения водохранилища ГЭС за интервал времени τ определяется уравнением (см. рис. 1)

$$\frac{\Delta V_i(t)}{\tau} = P_i(t) - Q_i(t) + Q_{i-1}(t), i = 1, \dots, N_m. \quad (1)$$

где i – номер конкретной ГЭС по течению реки в каскаде; N_m - количество ГЭС в каскаде m ; $V_i(t)$ – объем водохранилища в момент времени t , км³; $P_i(t)$ – полный или боковой приток воды в водохранилище, м³/с; $Q_i(t)$ – расход через створ ГЭС, м³/с; $Q_{i-1}(t)$ – расход через створ вышележащей ГЭС в каскаде, м³/с и, очевидно, $Q_0(t) = 0$.

Будем считать, что каскады ГЭС работают независимо друг от друга. Тогда параметры всех ГЭС во всех каскадах количеством M можно представить в виде векторов совокупной длиной $N = \sum_{m=1}^M N_m$, индекс $i = 1, \dots, N$ далее по тексту – это уникальный номер ГЭС. Номера ГЭС упорядочены по номеру каскада и расположению ГЭС по течению реки.

Для заданного периода времени $[t_0, t_T]$, разбитого на конечные интервалы продолжительностью $\tau_k = [t_{k-1}, t_k]$, $k = 1, \dots, T$, где t_0 – начальный период, T – количество

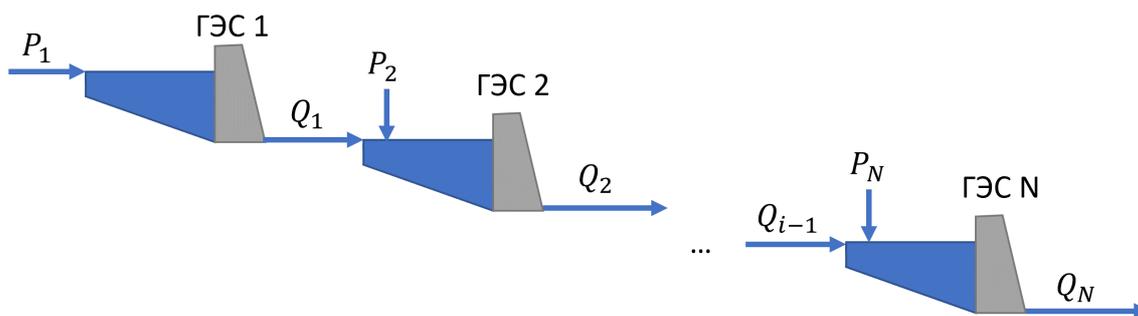


Рисунок 1: Гидрологическая модель каскада ГЭС.

интервалов, необходимо найти такие оптимальные расходы через ГЭС $Q_i(\tau_k)$, $i = 1, \dots, N$, которые удовлетворяют всем водохозяйственным, энергетическим, техническим ограничениям (включая характеристики ГЭС и для каждого установленного на ней гидроагрегата) на каждом интервале времени τ_k , позволяющие достигнуть максимальный совокупный доход при продаже выработанной электроэнергии на ГЭС, принадлежащих целевой компании при заданных (определенных) параметрах:

- $P_i(\tau_k) = P_{k,i}$ – прогнозных притоках воды в водохранилища ГЭС, m^3/c ;
- $Z_i(t_0) = Z_{0,i}$ – уровнях водохранилищ на начальный момент времени, m ;
- $Z_i(t_T) \geq Z_{T,i}$ – уровнях водохранилищ на конечный момент времени, m ;
- $c(\tau_k, E(\tau_k)) = c_k(\hat{E}_k)$ – цена электрической энергии на оптовом рынке, которая зависит также от суммарной выработанной электроэнергии \hat{E}_k всех ГЭС и других станциях в рассматриваемой энергосистеме за период τ_k , $руб. / кВт \cdot ч$;
- \tilde{E}_k – объем выработанной электроэнергии, $кВт \cdot ч$.

Здесь и далее для удобства восприятия нижний индекс k будет означать значение параметра в периоде τ_k .

Объем водохранилища ГЭС $V_i(t)$ на определенный момент времени t связан с уровнем водохранилища $Z_i(t)$ через эмпирическую зависимость и вычисляется через интерполяцию табличных данных.

Зависимость цены от суммарной выработки электроэнергии на оптовом рынке для каждого периода τ_k описывается линейно-кусочной функцией, пример которой приведен на рисунке 2. В данной работе предполагается, что такая линейно-кусочная функция определена на положительном множестве чисел, неотрицательна, непрерывна и не возрастает на всем множестве значений выработки электроэнергии (выработка тоже всегда больше нуля), т.е.

$$c_k(\hat{E}_k) \geq 0 \text{ для } 0 \leq E_k^{min} \leq \hat{E}_k \leq E_k^{max},$$

$$c_k(\hat{E}_k^1) \geq c_k(\hat{E}_k^2) \text{ для любых } \hat{E}_k^1 \leq \hat{E}_k^2. \quad (2)$$

Выработка электроэнергии для расчета цены на оптовом рынке за k -й период времени – это сумма выработки по всем ГЭС и остальной генерации \tilde{E}_k (например, совокупная выработка на ТЭЦ), которая исходно задается:

$$\hat{E}_k = \sum_{i=1}^N E_{k,i} + \tilde{E}_k.$$

Выработка электроэнергии i -й ГЭС в каскаде определяется формулой

$$E_{k,i} = 9,81 \cdot \eta_{k,i} \cdot q_{k,i} \cdot h_{k,i} \cdot \tau_k, \quad (3)$$

где $\eta_{k,i}$, $q_{k,i}$, $h_{k,i}$ – соответственно средние показатели КПД, суммарный расход через гидроагрегаты ГЭС, напор нетто ГЭС на указанном интервале времени. Расчет выработки электроэнергии за k -й период времени τ_k производится в следующем порядке:

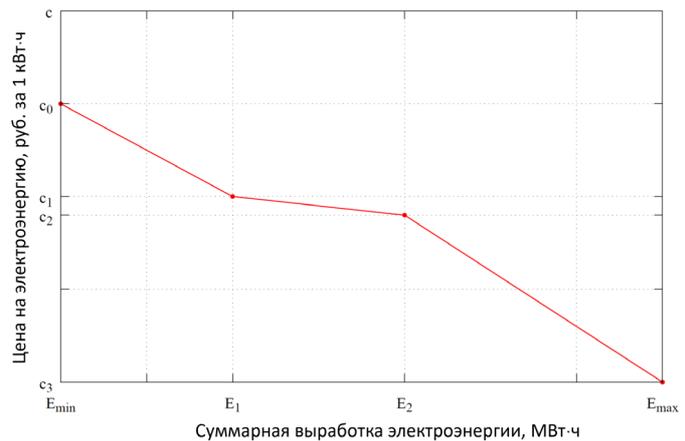


Рисунок 2: Пример зависимости цены электроэнергии от объема выработки в виде кусочно-линейной функции, которая задается для каждого периода k .

1. Рассчитывается средний напор нетто на основе разности значений средних уровней водохранилища $Z_i(t)$ и уровня нижнего бьефа $z_i(t)$ с использованием поправочных эмпирических коэффициентов α_i, β_i .

$$h_{k,i} = \left(\frac{(Z_{k,i} + Z_{k-1,i})}{2} - \frac{(z_{k,i} + z_{k-1,i})}{2} \right) \cdot \alpha_i + \beta_i.$$

2. Затем на основании вычисленного напора нетто рассчитываются расходы $Q_{k,i,j}$ через общее количество гидроагрегатов ГЭС равное GA_i таким образом, чтобы обеспечить максимальный КПД ГЭС, который равен среднему взвешенному КПД всех задействованных гидроагрегатов. КПД гидроагрегата определяется на основе проектных эксплуатационных характеристик гидроагрегата и в общем виде вычисляется через интерполяцию заданной двумерной нелинейной функции полиномиального вида, зависящей от расхода воды через гидроагрегат и напора нетто ГЭС. При этом необходимо также учитывать, что некоторые гидроагрегаты могут находиться в ремонте в расчетном периоде τ_k какое-то количество дней в соответствии с заданным планом-графиком ремонтов основного оборудования ГЭС. Т.е. необходимо решить задачу на максимум среднего КПД ГЭС

$$\max_{q_{k,i,j}} \eta_{k,i} = \max_{q_{k,i,j}} \sum_{j=1}^{GA_i} \eta_{k,i}(h_{k,i}, q_{k,i,j} \cdot r_{k,i,j}) \cdot q_{k,i,j} / q_{k,i} \quad (4)$$

при

$$0 \leq q_{k,i,j} \cdot r_{k,i,j} \leq q_{i,j}^{max} \cdot r_{k,i,j},$$

$$q_{k,i} = \sum_{j=1}^{GA_i} q_{k,i,j} \leq q_{i,j}^{max} \leq Q_{k,i},$$

где j – индекс гидроагрегата ГЭС; $r_{k,i,j}$ – коэффициент использования гидроагрегата в периоде τ_k от 0 до 1 в соответствии с исходно заданным планом-графиком ремонтов; причем

$$Q_{k,i} = q_{k,i} + s_{k,i}, \quad 0 \leq q_{k,i} \leq q_{k,i}^{max}, \quad (5)$$

где $s_{k,i}$ – холостой расход i -й ГЭС в периоде k , $q_{k,i}^{max}$ – максимально возможный расход через гидроагрегаты ГЭС.

Тогда выработка ГЭС при найденных в ходе решения задачи (4) КПД $\hat{\eta}_{k,i,j}$ и расходах гидроагрегатов $\hat{q}_{k,i,j}$ равна

$$E_{k,i} = 9,81 \cdot h_{k,i} \cdot \sum_{j=1}^{N_i} \hat{\eta}_{k,i,j} \cdot \hat{q}_{k,i,j} \cdot \tau_k. \quad (6)$$

Целевая функция в задаче максимизации дохода генерирующей компании, которой принадлежат только часть ГЭС из рассматриваемых каскадов может быть выражена как

$$\max_{Q_{i,k}} \sum_{k=1}^T \sum_{i \in I^*} E_{k,i} \cdot c_k(\hat{E}_k), \quad (7)$$

где оптимизируемые переменные - управляющие воздействия по всем периодам и всем ГЭС целевой компании (I^* - множество индексов таких ГЭС). Максимизация дохода выполняется при следующих основных условиях:

$$\text{на общие средние расходы ГЭС} \quad 0 \leq \underline{Q}_{k,i} \leq Q_{k,i} \leq \bar{Q}_{k,i}, \quad (8)$$

$$\text{на уровень водохранилища} \quad 0 \leq \underline{Z}_{k,i} \leq Z_{k,i} \leq \bar{Z}_{k,i}, \quad (9)$$

$$\text{на холостой расход} \quad 0 \leq s_{k,i} \leq \bar{s}_{k,i}, \quad (10)$$

$$\text{на изменение уровня водохранилища} \quad \underline{\Delta Z}_{k,i} \leq Z_{k,i} - Z_{k-1,i} \leq \bar{\Delta Z}_{k,i}, \quad (11)$$

$$\text{на изменение средних расходов ГЭС} \quad \underline{\Delta Q}_{k,i} \leq Q_{k,i} - Q_{k-1,i} \leq \bar{\Delta Q}_{k,i}, \quad (12)$$

а также при дополнительных ограничениях на диапазоны изменения уровня нижнего бьефа и напора нетто ГЭС.

3. Поиск оптимальных режимов работы каскадов ГЭС для обеспечения максимального дохода компании

Исходя из модели, описанной в главе 2, задача (7) – (14) - это задача нелинейной оптимизации вектора расходов ГЭС с набором ограничений на ключевые параметры режимов работы: общий расход, уровень водохранилища, уровень нижнего бьефа, напор нетто, холостой расход, изменение уровня, изменение расхода. В данной работе предложен подход, основанный на последовательном решении задач линейного программирования. Для этого, на основании анализа свойств модели и входящих в нее зависимостей, а также на базе реальной практики управления режимами ГЭС, введем следующее положение: поскольку выработка электроэнергии на ГЭС напрямую зависит от расхода воды через гидроагрегаты и напора нетто, то с увеличением расхода через гидроагрегаты ГЭС растет объем выработанной станцией электрической энергии (до появления холостых сбросов). Тогда процесс поиска режимов работы для решения оптимизационной задачи (7) может быть упрощен и выполнен в несколько этапов.

В качестве оптимизируемых переменных будем рассматривать компоненты вектора расходов $x = [q_{1,1}, q_{2,1}, \dots, q_{T,N}, s_{1,1}, s_{2,1}, \dots, s_{T,N}]$ из соотношения (5): расходы через гидроагрегаты и холостые расходы (сбросы) ГЭС. Тогда длина вектора x равна $(2 \cdot N \cdot T)$.

Этап 1. Поиск допустимого начального приближения.

Пусть задана начальная точка x_0 равная верхнему ограничению (8) на общие расходы ГЭС $x_0 = \bar{Q}$, которая в общем случае является недопустимой.

Тогда поиск допустимого начального приближения x_1 выполняется через решение задачи линейного программирования

$$\max_x \sum_{k=1}^T \sum_{i=1}^N q_{k,i} \cdot e_{k,i}(x_0) \quad (13)$$

где $e_{k,i} = E_{k,i}/(q_{k,i} \cdot \tau_k)$ — это удельная выработка электроэнергии ГЭС на объем сработанного водного ресурса водохранилища, используется как дополнительный весовой коэффициент. В данной постановке задачи будем использовать основные ограничения (8) - (12). Ограничение на уровень водохранилища (9) можно переписать через уравнение водного баланса в виде ограничения на объем водохранилища в начале и в конце интервал времени τ_k :

$$\underline{V}_{k,i} \leq V_{k,i} = V_{k-1,i} + (P_{k,i} - Q_{k,i} + Q_{k,i-1}) \cdot \tau_k \leq \bar{V}_{k,i}, \quad (14)$$

так как уровень и объем водохранилища связаны через эмпирическую монотонную зависимость $V_{k,i} = v(Z_{k,i})$, $k = 0, \dots, T$.

В ходе решения задачи (13) с ограничениями (8) - (12), (14) получаем допустимую точку x_1 , которая задает режимы работы ГЭС в каскадах, которые обеспечивают максимум выработки электрической энергии за заданный период.

Этап 2. Поиск оптимального решения методом градиентного спуска.

Начиная с найденной на предыдущем этапе допустимой точки x_1 , итеративно уточняем полученное решение

$$x_{n+1} = x_n + \lambda \cdot \nabla f(x_n). \quad (15)$$

Величина шага λ по направлению градиента целевой функции f (функция дохода генерирующей компании), определяется с помощью решения следующей задачи линейного программирования:

$$\max_{Q_{k,i}} \sum_{k=1}^T \sum_{i=1}^N |Q_{k,i} - Q_{k,i}^{x_n}| \cdot \frac{\nabla f(x_n)}{\|\nabla f(x_n)\|}, |Q_{k,i} - Q_{k,i}^{x_n}| < \varepsilon. \quad (16)$$

где $Q_{k,i}^{x_n}$ — общие расходы через створ ГЭС, полученные на предыдущей итерации; при этом используются те же ограничения и зависимости балансовых уравнений (8) - (12), (14); поиск новых расходов $Q_{k,i}$ производится в окрестности текущей точки $[x_n \pm \varepsilon]$, где ε — некоторая заданная точность расчета среднемесячных расходов ГЭС.

Полученное с помощью метода градиентного спуска на этапе 2 решение x^* может быть дополнительно улучшено в его окрестности ε перебором каскадных возмущений общих расходов через ГЭС.

4. Реализация и предварительные результаты

Для решения задач линейного программирования (13) и (16) использован свободно доступный математический пакет OR-Tools (набор решений Google с открытым исходным кодом и интерфейсом расширений) с использованием GLOP-алгоритма на основе симплекс-метода [14]. Общая схема решения задач по оптимальному планированию режимов состоит из 4 этапов:

1. Экспресс-оценка непротиворечивости заданных режимных ограничений (8) – (12) и подтверждение существования области допустимых значений.
2. Поиск начальной точки внутри допустимой области через решение задачи линейного программирования (13) – (14).
3. Если допустимая точка найдена, то ищем максимум функции дохода генерирующей компании градиентным методом (15) – (16).
4. Проверка возможности покоординатного улучшения полученного решения на этапе 3 перебором каскадных возмущений в окрестности точки решения.

Описанный в предыдущем разделе двухэтапный алгоритм реализован в составе программно-вычислительного комплекса с подсистемой задания необходимых исходных данных и визуализацией результатов полученного решения. Проведена апробация расчетной модели и оптимизационного алгоритма на примере статистически усредненных данных Ангаро-Енисейского каскада ГЭС. Ангарский каскад состоит из 4 ГЭС: Иркутская, Братская, Усть-Илимская, Богучанская, три первых из которых принадлежат целевой генерирующей компании. Енисейский каскад состоит из 3 ГЭС: Саяно-Шушенская, Майнская, а также Красноярская, которая также принадлежит целевой генерирующей компании. Рассматривался период $T = 13$ месяцев. Заданы произвольные линейно-кусочные функции цен на электроэнергию по 4 основным сезонам вида как на рис. 2. Результаты оптимизационных расчетов показывают возможность повышения финансово-экономической эффективности генерирующей компании на 5% по сравнению с управлением режимами на основе средних фактических данных.

5. Заключение

В работе описана математическая модель режимов работы ГЭС в каскаде, содержащего ГЭС генерирующей компании, которой необходимо получить максимальную прибыль при генерации электроэнергии на ГЭС с учетом цен на оптовом рынке. Модель позволяет осуществлять расчет дохода генерирующей компании на основе прогнозных ценовых кривых. Сделана математическая постановка задачи максимизации дохода генерирующей компании на прогнозный период. Предложен алгоритм поиска оптимальных режимов работы ГЭС для решения задачи максимизации дохода генерирующей компании. Разработан прикладной программный продукт, реализующий вышеуказанную расчетную модель и алгоритм оптимизации.

6. Список литературы

- [1] В. В. Колмогоров, Л. Е. Халяпин, Иркутская ГЭС и Ангаро-Енисейский каскад ГЭС в энергетике Сибири. Пути повышения эффективности, *ЕСО*, т. 52, вып. 8, сс. 44–53, июл. 2022, doi: [10.30680/ЕСО0131-7652-2022-8-44-53](https://doi.org/10.30680/ЕСО0131-7652-2022-8-44-53).
- [2] He, Z., Zhou, J., Xie, M., Jia, B., Qin, H., Zhang, H., Study on guaranteed output constraints in the long term joint optimal scheduling for the hydropower station group, *Energy*. (2019) doi:[10.1016/j.energy.2019.07.066](https://doi.org/10.1016/j.energy.2019.07.066)

- [3] S. Zhang, M. Zhou, and T. Hu, An optimization model for joint power generation dispatching of the cascade hydropower stations on the lower reaches of Jinsha river and the Yangtze river, in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing Ltd, Mar. 2021. doi: [10.1088/1755-1315/676/1/012122](https://doi.org/10.1088/1755-1315/676/1/012122).
- [4] J. Chen, Long-Term Joint Operation of Cascade Reservoirs Using Enhanced Progressive Optimality Algorithm and Dynamic Programming Hybrid Approach, *Water Resources Management*, vol. 35, no. 7, pp. 2265–2279, May 2021, doi: [10.1007/s11269-021-02846-y](https://doi.org/10.1007/s11269-021-02846-y).
- [5] C. Ji, Z. Jiang, P. Sun, Y. Zhang, and L. Wang, Research and Application of Multidimensional Dynamic Programming in Cascade Reservoirs Based on Multilayer Nested Structure, 2014, doi: [10.1061/\(ASCE\)WR.1943-5452](https://doi.org/10.1061/(ASCE)WR.1943-5452).
- [6] Feng, Zhong-kai, Wen-jing Niu, and Chun-tian Cheng, Optimizing Electrical Power Production of Hydropower System by Uniform Progressive Optimality Algorithm Based on Two-Stage Search Mechanism and Uniform Design, *Journal of Cleaner Production*, vol. 190, (2018): 432–42. doi: [10.1016/j.jclepro.2018.04.134](https://doi.org/10.1016/j.jclepro.2018.04.134).
- [7] T. Zhao and J. Zhao, Improved multiple-objective dynamic programming model for reservoir operation optimization, *Journal of Hydroinformatics*, vol. 16, no. 5, pp. 1142–1157, 2014, doi: [10.2166/hydro.2014.004](https://doi.org/10.2166/hydro.2014.004).
- [8] Shen, Jianjian, Xiufei Zhang, Jian Wang, Rui Cao, Sen Wang, and Jun Zhang, Optimal Operation of Interprovincial Hydropower System Including Xiluodu and Local Plants in Multiple Recipient Regions, *Energies*, vol. 12, no. 1 (2019): 144. doi: [10.3390/en12010144](https://doi.org/10.3390/en12010144).
- [9] Wang, J., Cheng, C., Shen, J., Cao, R., & Yeh, W. W.-G., Optimization of Large-Scale Daily Hydrothermal System Operations With Multiple Objectives, *Water Resources Research*, 54(4), 2834–2850. (2018) doi:[10.1002/2017wr021291](https://doi.org/10.1002/2017wr021291)
- [10] X. Ma, H. Hu, and Y. Ren, “A hybrid deep learning model based on feature capture of water level influencing factors and prediction error correction for water level prediction of cascade hydropower stations under multiple time scales,” *Journal of Hydrology*, vol. 617, Feb. 2023, doi: [10.1016/j.jhydrol.2022.129044](https://doi.org/10.1016/j.jhydrol.2022.129044).
- [11] Бердников В.М., Абасов Н.В., Осипчук Е.Н. Применение нейросетевых методов для формирования долгосрочных режимов ГЭС, Сборник тезисов IV Международной конференции “Системные исследования Системные исследования в энергетике 2023”. 2023. С.68-69.
- [12] J. J. Shen, H. Zhong, X. Cheng, A Nonlinear Method for Long-term Optimal Operations for Cascade Hydropower Plants with Electricity-Price Curves, *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, June 2020. doi: [10.1088/1755-1315/495/1/012023](https://doi.org/10.1088/1755-1315/495/1/012023).
- [13] Abasov N.V., Nikitin V.M., Osipchuk E.N. System Of Models To Study Long-Term Operation Of Hydropower Plants In The Angara Cascade, *Energy Systems Research*, Vol. 2, No.2, 2019. P.5-18. doi: [10.25729/esr.2019.02.0001](https://doi.org/10.25729/esr.2019.02.0001)
- [14] L. Perron, F. Didier, и S. Gay, The CP-SAT-LP Solver, in: *29th International Conference on Principles and Practice of Constraint Programming (CP 2023)*, eds. R.H.C.Yap, Leibniz International Proceedings in Informatics (LIPIcs), vol. 280, Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, pp. 3:1-3:2. doi: [10.4230/LIPIcs.CP.2023.3](https://doi.org/10.4230/LIPIcs.CP.2023.3)

Актуальные вопросы оценки аспектов информационной безопасности в модели интероперабельности организационно-технических систем

Алексей Нестеров ¹

¹ Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, набережная реки Мойки, д. 61, литера А, 191186, Санкт-Петербург, Россия

Аннотация

В современных условиях при интеграции организационно-технических систем (ОТС) возрастает актуальность обеспечения интероперабельности в таких системах. В докладе проведен обзор работ по тематике интероперабельности. Рассмотрен ряд работ, выполненных в области параметров информационной безопасности, как одного из аспектов технической интероперабельности. Выявлена проблемная ситуация - необходимость повышения качества интероперабельности на техническом уровне при взаимодействии ОТС в условиях отсутствия научно-методического аппарата количественной оценки параметров информационной безопасности. В докладе научная задача сформулирована как необходимость разработки элементов научно-методического аппарата количественной оценки параметров информационной безопасности и повышения интероперабельности при взаимодействии ОТС.

Ключевые слова

интероперабельность, технический уровень интероперабельности, организационно-техническая система, параметры информационной безопасности.

1. Введение

В настоящее время актуальным вопросом развития различных автоматизированных систем и информационных систем является их взаимодействие и сопряжение между собой. Взаимодействие и сопряжение различных автоматизированных систем и информационных систем актуализирует вопросы обеспечения интероперабельности таких систем. Интероперабельность - способность двух или более информационных систем или компонентов к обмену информацией и к использованию информации, полученной в результате обмена.

Значимый вклад в вопросы обеспечения интероперабельности в Российской Федерации внесли исследования ученых научной школы, возглавляемой сотрудником Института радиотехники и электроники им. В.А. Котельникова Российской академии наук профессором А.Я. Олейниковым. В ходе этих исследований получены значимые частные результаты в области интероперабельности для информационных систем самого широкого класса и назначения – в здравоохранении [1-2], в электронной коммерции [3], в образовательном процессе [4], в научных исследованиях [5], в облачных вычислениях [6], в промышленности [7], а также в области обеспечения обороноспособности нашей страны [8-11]. Различные аспекты проблемы интероперабельности исследуются специалистами в других областях, например, в области здравоохранения – В.А. Дрогвозом [12-14]. С.В. Козлов в своей работе [15] рассматривает вопросы использования процессного подхода к обеспечению

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: nesterov.aa@sut.ru (A. 1);

ORCID: 0009-0006-0793-9549 (A. 1);



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.20

интероперабельности функциональных систем при создании интегрированных систем управления, в том числе для обоснования практических мер по парированию негативного влияния внешних факторов. В настоящее время исследования интероперабельности представителями научной школы А.Я. Олейникова продолжаются, и среди значимых результатов можно выделить работы [16-22]. В этой связи следует отметить работу [23], в которой С.И. Макаренко представлена расширенная описательная модель интероперабельности организационно-технических систем.

В соответствии с эталонной моделью взаимодействие информационных систем формализуется на трех иерархических уровнях интероперабельности: техническом, семантическом, организационном. Технический уровень интероперабельности - уровень интероперабельности, на котором формализуются процессы информационного взаимодействия между техническими системами, техническими средствами, аппаратными и программными комплексами с учетом особенностей реализации их интерфейсов и протоколов обмена информацией, а также форм и форматов представления информации [21]. Аспекты и параметры, составляющие содержание технического уровня интероперабельности, представлены на рис. 1. Одним из аспектов технического уровня интероперабельности являются параметры информационной безопасности.



Рисунок 1: Аспекты и параметры, составляющие содержание технического уровня интероперабельности

2. Параметры информационной безопасности в модели интероперабельности организационно-технических систем

Говоря о параметрах информационной безопасности, необходимо учитывать данное в [23] определение: «Информационная безопасность (ИБ) – это состояние, при котором обеспечивается конфиденциальность, целостность и доступность информации». При этом для развития и формализации параметров ИБ отечественной модели интероперабельности предлагается выделить три ключевые области:

1. Конфиденциальность – параметр, гарантирующий соблюдение ролевой модели доступа к информации.
2. Целостность – параметр, гарантирующий соблюдение ролевой модели при попытке изменения данных.
3. Доступность – параметр, гарантирующий факт наличия доступа в рамках ролевой модели.

Рассматривая свойство информационной безопасности как систему, состоящую из трех последовательно включенных свойств: конфиденциальности, целостности и доступности,

можно принять, что недостижение заданного уровня обеспечения для хотя бы одного из трех свойств информационной безопасности приводит к пропорциональному снижению общего уровня защищенности системы. Таким образом, простейшая модель информационной безопасности состоит из трех последовательно включенных составляющих и представлена на рис. 2:



Рисунок 2: Простейшая модель информационной безопасности

В теории ИБ рассматриваются вопросы обеспечения целостности, достоверности и конфиденциальности в условиях воздействия внешних дестабилизирующих факторов – информационно-технических воздействий (ИТВ). Как следствие, при рассмотрении аспектов обеспечения интероперабельности организационно-технических систем (ОТС) необходимо учитывать возможность ИТВ злоумышленников. Все ИТВ можно разделить на совместные и несовместные. Совместные ИТВ – это ИТВ, которые могут производиться одновременно, например, воздействие компьютерными вирусами и распределенное отправление на атакуемый сервер запросов с большого количества устройств (DDoS-атака). В свою очередь, несовместные ИТВ – это ИТВ, которые не могут производиться одновременно, например, проведение сетевой разведки и атака по выявленным в результате этой разведки уязвимостям системы (рис. 3).

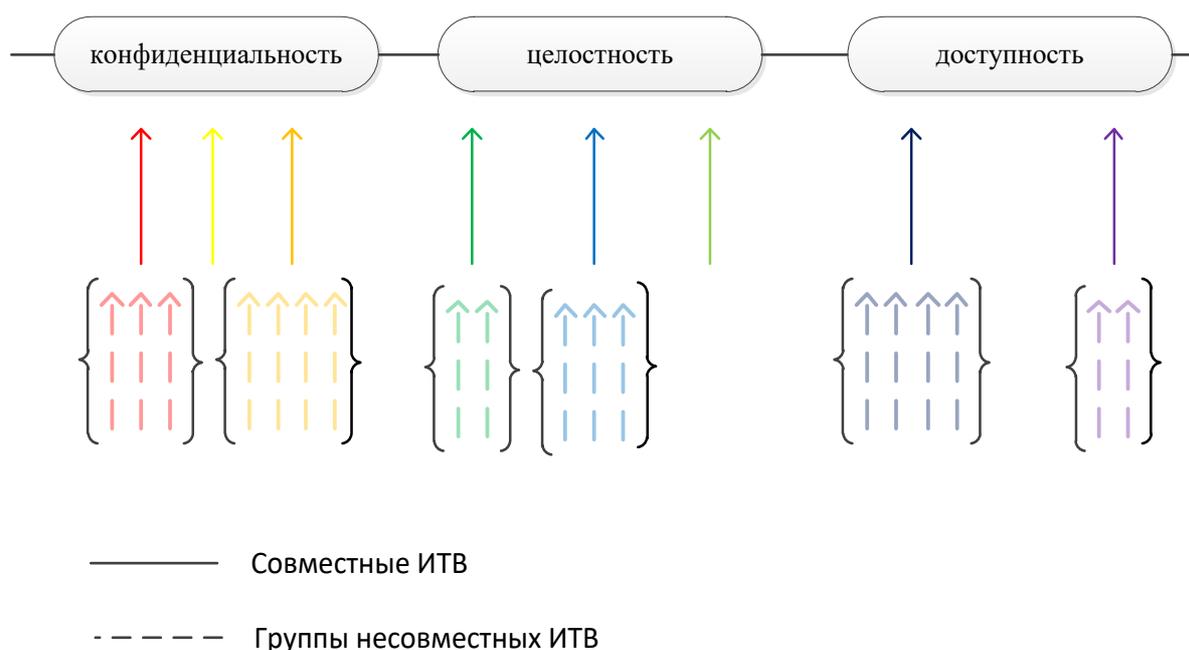


Рисунок 3: Совместные и несовместные ИТВ на свойства информационной безопасности

В [23] показано, что количественная оценка свойств информационной безопасности в рамках оценки технической интероперабельности ОТС зависит от параметров информационной безопасности, однако при этом научно-методического аппарата для количественной оценки параметров информационной безопасности автором не предлагалось. Аспекты ИБ в контексте интероперабельности рассмотрены в работах [24-26], однако в этих работах отсутствует их формализация. В работах [27-28] в достаточной степени проработана оценка рисков ИБ применительно к системам различного типа, однако, вопросы ИБ применительно к интероперабельности не рассмотрены.

Вышеуказанные факторы позволили сформулировать проблемную ситуацию как необходимость повышения качества интероперабельности на техническом уровне при взаимодействии ОТС в условиях отсутствия научно-методического аппарата количественной

оценки параметров информационной безопасности и повышения интероперабельности в составе теории систем. Для разрешения данной проблемной ситуации может быть сформулирована актуальная цель исследования – повышение качества интероперабельности на техническом уровне при взаимодействии ОТС за счет разработки элементов научно-методического аппарата количественной оценки параметров информационной безопасности. Объектом исследования являются параметры информационной безопасности взаимодействующих ОТС, а предметом исследования – интероперабельность взаимодействующих ОТС. Научной задачей является разработка элементов научно-методического аппарата количественной оценки параметров информационной безопасности и повышения интероперабельности при взаимодействии ОТС.

Решением этой научной задачи будет научный результат, значимый для развития интероперабельности ОТС и обладающий практическим эффектом в части повышения технической интероперабельности путем применения элементов разработанного научно-методического аппарата.

3. Список литературы

- [1] Гончаров Н. Г., Гулиев Я. И. О., Гуляев Ю. В., Кавинская Ю. М., Каменщиков А. А., Олейников А. Я., Хаткевич М. И. Вопросы создания единого информационного пространства в системе здравоохранения РАН // Информационные технологии и вычислительные системы. 2006. № 4. С. 83-95.
- [2] Анциперов В. Е., Каменщиков А. А., Кочуков А., Никитов Д. С., Олейников А. Я. Функциональная стандартизация при создании медицинских информационных систем // Врач и информационные технологии. 2006. № 4. С. 15-18.
- [3] Олейников А. Я., Разинкин Е. И. Профиль интероперабельности в области электронной коммерции // Информационные технологии и вычислительные системы. 2013. № 4. С. 74-79.
- [4] Олейников А. Я., Меркулова А. В. К вопросу о построении интегрированной корпоративной информационной среды вуза // Журнал радиоэлектроники. 2005. № 5. С. 4. – URL: <http://jre.cplire.ru/iso/nov05/2/text.html> (дата доступа: 24.05.2024).
- [5] Батоврин В. К., Васютович В. В., Журавлев Е. Е., Олейников А. Я., Петров А. Б., Соколов С. А., Теряев Е. Д. Построение профиля информационных, вычислительных и телекоммуникационных ресурсов для обеспечения фундаментальных исследований // Журнал радиоэлектроники. 2001. № 11. С. 8. – URL: <http://jre.cplire.ru/koi/dec01/8/text.html> (дата доступа: 24.05.2024).
- [6] Журавлев Е. Е., Иванов С. В., Каменщиков А. А., Олейников А. Я., Разинкин Е. И., Рубан К. А. Интероперабельность в облачных вычислениях // Журнал радиоэлектроники. 2013. № 9. С. 14.
- [7] Олейников А. Я., Егоров Г. А., Журавлев Е. Е., Королев А. С., Кочуков А. Н., Широбокова Т. Д. Применение технологии открытых систем для создания интегрированных информационных систем промышленных предприятий // Радиопромышленность. 2006. № 2. С. 90-107.
- [8] Башлыкова А. А., Каменщиков А. А., Олейников А. Я. О подходах к разработке профилей интероперабельности в военной области // Информационные технологии и вычислительные системы. 2017. № 4. С. 112-121.
- [9] Каменщиков А. А., Олейников А. Я., Чусов И. И., Широбокова Т. Д. Проблема интероперабельности в информационных системах военного назначения // Журнал радиоэлектроники. 2016. № 11. С. 16.
- [10] Башлыкова А. А., Каменщиков А. А., Олейников А. Я. Обеспечение интероперабельности как средства бесшовной интеграции функциональных подсистем в составе перспективных автоматизированных систем военного назначения // Журнал радиоэлектроники. 2018. № 9. С. 18.

- [11] Быстров Р. П., Корниенко В. Н., Олейников А. Я. Интероперабельность, информационное противоборство и радиоэлектронная борьба // Успехи современной радиоэлектроники. 2018. № 5. С. 15-34.
- [12] Дроговоз В.А. Совершенствование оценки показателей интероперабельности сложных систем при помощи сценарного моделирования // Сборник трудов XXIX Международной научно-технической конференции «Радиолокация, навигация, связь», посвященной 70-летию кафедры радиофизики ВГУ (г. Воронеж, 18–20 апреля 2023 г.). Том 5. – Воронеж: ВГУ, 2023. – С. 361-369.
- [13] Дроговоз В.А. Обеспечение интероперабельности электронного здравоохранения в условиях цифровой трансформации экономики России // Сборник трудов XII Международной научной конференции «ИТ-Стандарт 2023». – М.: Издательство «Перспект», 2023. – С. 89-97.
- [14] Дроговоз В.А. Концептуальные аспекты обеспечения интероперабельности электронного здравоохранения в условиях цифровой трансформации экономики России // ИТ-стандарт. 2023. №2. С. 51-61.
- [15] Козлов С.В. Процессные аспекты интероперабельности интегрированных систем управления // ИТ-Стандарт. 2019. № 1(18). С. 25–30.
- [16] Олейников А.Я., Растягаев Д.В., Фомин И.А. Основные положения концепции обеспечения интероперабельности сетевых информационных систем // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. 2020. № 3. С. 122–131. DOI:10.25586/RNU.V9187.20.03.P.122
- [17] Олейников А.Я. Актуальное состояние проблемы интероперабельности // ИТ-Стандарт. 2020. № 2(23). С. 37–42.
- [18] Башлыкова А.А., Зацаринный А.А., Каменщиков А.А., Козлов С.В., Олейников А.Я., Чусов И.И. Интероперабельность как научно-методическая и нормативная основа бесшовной интеграции информационно-телекоммуникационных систем // Системы и средства информатики. 2018. Т. 28. № 4. С. 61–72. DOI:10.14357/08696527180407
- [19] Башлыкова А.А., Гаджикулиев Т.А., Олейников А.Я. Решение проблемы интероперабельности в проектах «умного города» // Современные информационные технологии и ИТ-образование. 2019. Т. 15. № 3. С. 767–774. DOI:10.25559/SITITO.15.201903.767-774
- [20] Макаренко С.И. Интероперабельность человеко-машинных интерфейсов. СПб.: Научно-технологические технологии, 2023. 185 с.
- [21] Макаренко С.И., Нестеров А.А. Структурно-функциональная модель интероперабельности организационно-технических систем. // Труды учебных заведений связи. 2023. – Т. 9, № 4. – С. 65-74.
- [22] Нестеров А.А. Нормативно-правовая база для обеспечения интероперабельности организационно-технических систем. // Сборник трудов XII Международной научной конференции «ИТ-Стандарт 2023» - М.: Издательство «Перспект», 2023. – С. 113-120.
- [23] Макаренко, С. И. Интероперабельность организационно-технических систем / С. И. Макаренко. – Санкт-Петербург : Научно-технологические технологии, 2024. – 313 с.
- [24] Т. Е. Черницкая, С. И. Макаренко, Д. В. Растягаев. Аспекты информационной безопасности в рамках оценки интероперабельности сетевых информационных систем // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. – 2020. – № 4. – С. 113-121.
- [25] Макаренко С. И., Олейников А. Я., Черницкая Т. Е. Модели интероперабельности информационных систем // Системы управления, связи и безопасности. 2019. № 4. С. 215-245. DOI: 10.24411/2410-9916-2019-10408
- [26] Олейников А.Я. Проблема интероперабельности в платформе Industry 4.0. и смежных областях. IX Международная конференция ИТ-Стандарт 2019, 2019.
- [27] Остапенко А. Г. Функция возможности в оценке рисков, шансов и эффективности систем // Информация и безопасность. 2010. Т. 13. № 1. С. 17-20.
- [28] Остапенко А. Г. Инновационные шансы и риски: эвентологическое развитие теории информации // Информация и безопасность. 2012. Т. 15. № 3. С. 313-320.

Сервис автоматизации создания интеллектуальных систем на платформе Яндекс.Облако

Ольга А. Николаичук¹, Александр И. Павлов¹

¹ *Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук (ИДСТУ СО РАН), ул. Лермонтова 134, Иркутск, Россия*

Аннотация

В работе проведён анализ возможности использования инфраструктуры Яндекс.Облако для размещения разрабатываемой авторами платформы создания систем, основанных на знаниях. Рассмотрен вопрос реализации пользовательского интерфейса платформы и создаваемых с ее помощью систем. Приведено описание механизма реализации поведения системы, основанной на знаниях, на стороне клиента, а также перечислены использованные ограничения. Приведены примеры реализованных форм пользовательского интерфейса в предметной области надёжность механических систем. Осуществлен обзор вариантов реализации облачного сервиса на основе платформы, включая вопрос отчуждения копии, разработанной с помощью сервиса системы.

Ключевые слова

Системы, основанные на знаниях, AI Paas, Яндекс.Облако,

1. Введение

Одной из примет настоящего времени является значительный рост объема информации, как структурированной, так и не структурированной, которую необходимо обрабатывать в процессе выполнения различных исследовательских задач. При этом для целого ряда подобных задач характерно постоянное изменение требований, а также возможность повторного использования полученных результатов в другой предметной области (повторное использование кода). На текущий момент уже существует большое количество программных средств, с помощью которых могут быть решены подобные задачи, начиная от инструментальных средств общего назначения, до специализированных на конкретную задачу библиотек. Однако при этом многие из них обеспечивают решение на общем уровне, и требуют описания специфики решаемой задачи с использованием некоторого языка программирования, что является проблемой для специалиста-предметника (высокие требования, связанные со знанием языка программирования). В то же время, для исследователей, которые могут использовать существующий арсенал программных средств, имеется проблема повторного использования полученных результатов, когда для их повторного использования в другой предметной области требуется адаптация решения на основе знаний предметной области (проблема недостатка предметно-ориентированных знаний). Для решения указанных проблем авторами предложен специализированный инструмент — платформа разработки систем, основанные на знаниях (СОЗ) [1]. Отличием предложенной платформы от существующих инструментальных средств («HiAsm» [2] или «LabVIEW» [3] и т. п.) являются реализация визуального программирования на основе набора базовых операций (функций компонентов) и их комбинаций (композиционных операций). При этом композиционные операции являются средством обеспечения доступности результатов

16th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: nikoly@icc.ru (A. 1); asd@icc.ru (A. 2);

ORCID: 0000-0002-5186-0073 (A. 1); 0000-0002-7753-7514 (A. 2);

DOI: 10.47350/ICCS-DE.2024.21

(методов и их программной реализации) для повторного использования за счет возможности добавления их к базовому набору.

Предложенная платформа обеспечивает возможность создания прикладной СОЗ с помощью визуального конструирования требуемого поведения, как императивного, так и декларативного с помощью визуальных редакторов компонентов платформы. Основными компонентами платформы являются: компонент работы с реляционными базами данных (БД) [1], компонент онтологического моделирования предметной области [4,5], компонент разработки продукционных баз знаний [6] и компонент формализации императивного поведения [7].

В данной работе обсуждаются задачи создания пользовательского интерфейса и реализации облачного сервиса на основе программной платформы для разработки СОЗ.

2. Реализация пользовательского интерфейса платформы

В общем случае разработка пользовательского интерфейса (UI) прикладной программной системы является трудоемкой задачей, решение которой может занимать до половины и более от общего количества времени, затраченного на разработку, с учетом практически неизбежного возникновения необходимости в различных изменениях и доработках. В рассматриваемой работе задача создания UI присутствует в следующих подзадачах: организация взаимодействия пользователей с платформой создания СОЗ, с входящими в ее состав компонентами, а также с разрабатываемыми на ее основе прикладными СОЗ. Высокая трудоемкость создания независимого UI при решении каждой из подзадач обуславливает необходимость единого подхода для всех перечисленных случаев. При этом достичь полной стандартизации не представляется возможным ввиду большой вариативности предметных задач. Поэтому в качестве цели для унификации UI выбрана функциональность взаимодействия с отдельными объектами данных, включая просмотр текущего состояния, создание нового, редактирование и удаление существующего; а также работы с массивами (списками) подобных объектов, включая просмотр и фильтрацию по параметрам. В качестве средства для визуального отображения при реализации выбранной функциональности предложено использовать табличное представление.

2.1. Пользовательский интерфейс: табличное представление

Представление информации в табличном виде уже используется в компоненте работы с базами данных [1], в котором имеется возможность выполнения стандартных операций, а также поиска информации для любой из таблиц заданной БД. В настоящее время существует два варианта UI компонента: (1) с использованием jQuery (библиотеки jQueryUI и jQueryGrid), (2) с использованием Vue.js (библиотека Quazar [8]). В качестве основы для реализации стандартного UI платформы, ее компонентов и целевых СОЗ был выбран 1-й вариант.

Компонент работы с базами данных использует спецификацию таблицы БД для создания UI, предназначенного для работы именно с этой таблицей. Для представления такого рода спецификаций используется модель объекта базы данных, которая на стороне сервера формализуется интерфейсом *IDescriptionOfDatabaseObject* (Рисунок 1), а на стороне клиента представляет собой JSON объект.

Механизм формирования UI компонента работы с базами данных предложено модифицировать следующим образом:

1. добавить возможность сформировать для произвольного объекта его спецификацию в формате, соответствующем *IDescriptionOfDatabaseObject*, при необходимости оставив без значения отсутствующие свойства;
2. добавить возможность представления состояния произвольного объекта в виде массива простых объектов вида: имя/значение, для решения задачи просмотра и редактирования отдельно взятого объекта.

Кроме того, существующий UI платформы и её компонентов предложено заменить на иерархическую структуру, где на верхнем уровне находятся функциональность платформы (представленная компонентами), а ниже полученные результаты, которые в свою очередь

подразделяются на составные части. Так функция «концептуальное моделирование» содержит перечень созданных моделей, которые в свою очередь делятся на понятия, свойства, рабочие области и т. д., а функция «создание производственных баз знаний» содержит перечень созданных баз знаний, которые в свою очередь делятся на шаблоны, правила, начальные условия.

```

▼ {Status: true, Data: [...], Message: "",...}
  AdditionalData: null
  Data: [...],
  ▼ Data: [...],
    ▶ 0: {ID: 2249, p1: "Деградационный процесс", p2: null, p3: [null, "Значение не присвоено"], p4: ["33",...],...}
    ▶ 1: {ID: 2248, p1: "Работа", p2: null, p3: [null, "Значение не присвоено"], p4: ["33",...], p5: false}
    ▶ 2: {ID: 2247, p1: "Дефекты (повреждения)", p2: null, p3: [null, "Значение не присвоено"], p4: ["33",...],...}
    ▶ 3: {ID: 2246, p1: "Среда функционирования", p2: null, p3: [null, "Значение не присвоено"], p4: ["33",...],...}
    ▶ 4: {ID: 2245, p1: "Технология изготовления", p2: null, p3: [null, "Значение не присвоено"], p4: ["33",...],...}
    ▶ Description: {StartIndex: 0, CountOfRecords: "1320", IndexOfPage: 1, SizeOfPage: 5, CountOfPages: 264}
    Message: ""
  ▼ MetaInformation: {NameOfInterface: "IDescriptionOfDatabaseObject", CompositeName: ["Концептуальные модели", "Понятия"],...}
    Caption: "\"Концептуальные модели\".\"Понятия\""
    CheckConstraints: [{NameOfInterface: "IDescriptionOfCheckConstraint",...}]
    CompositeName: ["Концептуальные модели", "Понятия"]
    Description: null
    HandlerOfSearchCondition: {InitialSetOfCompareOperators: ["eq", "ne"], SetOfNumericCompareOperators: ["lt", "le", "gt", "ge"],...}
    ID: "259068"
    Identifier: {NameOfInterface: "IDescriptionOfIdentifier", CompositeName: ["Концептуальные модели", "Понятия_pkey"],...}
    NameOfInterface: "IDescriptionOfDatabaseObject"
    ObjectCollectionProperties: [{NameOfInterface: "IDescriptionOfCollectionOfObjectsProperty",...},...]
    ObjectProperties: [{NameOfInterface: "IDescriptionOfObjectProperty",...},...]
    Properties: [...],
    UniqueIndices: [{NameOfInterface: "IDescriptionOfUniqueIndex",...}]
  Status: true

```

Рисунок 1: Пример входных данных компонент формирования пользовательского интерфейса в режиме работы с базой данных

Важную роль в предложенной архитектуре UI занимает концепция связанных с выделенной строкой действий, которая заключается в возможности добавить в описание иерархического уровня спецификацию действий, которые могут быть выполнены с выделенной строкой таблицы. Предложенная концепция позволяет в некоторой степени преодолеть относительно низкую выразительность представления UI в виде таблицы. Той же цели служит набор проблемно-ориентированных редакторов, функциональность которых не укладывается в ограничения табличного представления, в качестве примера можно назвать редактор правил или редактор композитных операций. В предложенной архитектуре UI редакторы такого рода реализуются в виде отдельных модулей и вызываются как действия, связанные с выделенной строкой.

Таким образом, для решения задачи обеспечения платформы унифицированным UI предложена концепция отображения иерархической структуры обрабатываемых данных от общего к частному в некоторой аналогии с работой файловых менеджеров, для реализации которой использован компонент формирования пользовательского интерфейса, в функциональность которого входит просмотр и модификация данных произвольных объектов на основе их описания (метаданных), а также набор специализированных проблемно-ориентированных редакторов оформленных в виде отдельных модулей.

На рисунке 2 показан пример пользовательского интерфейса компонента разработки производственных баз знаний, отображен экран просмотра производственных правил, включая перечень действий для выбранного правила, а также окно со специализированным редактором правил.

2.2. Реализация поведения функции СОЗ на стороне клиента

Процесс создания прикладной СОЗ с помощью предложенной платформы включает описание требуемого поведения в виде композитной операции с помощью компонента формализации императивного поведения. В ряде случаев выполнение композитной операции на стороне сервера может быть затруднено, например, когда время выполнения может превысить установленные ограничения, либо когда необходимо взаимодействие с пользователем в процессе выполнения. Поэтому предложено обеспечить возможность

выполнения композитной операции не только на стороне сервера, но и на стороне клиента. Для этого в компонент добавлена функциональность отображения заданной композитной операции в код на языке Javascript. Процесс генерации в целом осуществляется по аналогии с созданием кода на языке PHP, с учетом асинхронных особенностей языка Javascript. Важной особенностью выполнения композитных операций на стороне клиента является возможность ее реализации в виде набора обработчиков для некоторого списка событий, т. е. результатом выполнения кода, сгенерированного для операции, может быть подготовка и запуск инфраструктуры обработки Websocket событий. Данную особенность предлагается использовать для реализации поддержки пользователя в процессе выполнения прикладной задачи, когда пользователь может получать дополнительную информацию и указания в зависимости от состояния наблюдаемых параметров.

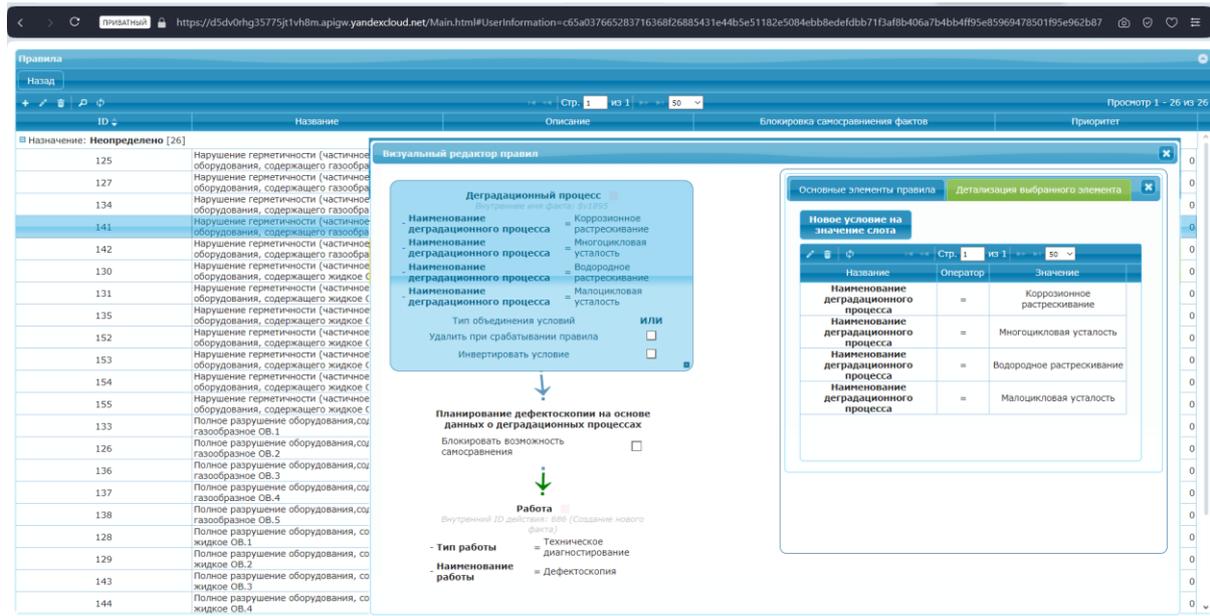


Рисунок 2: Пример UI компонента разработки продукционных баз знаний

Отметим, что в данный момент генерация Javascript кода ограничена возможностью выполнения команд на сервере, которая реализуется с помощью отправки на сервер либо HTTPS запроса, либо Websocket сообщения, а также возможностью отображения полученных в ответ результатах. При этом запуск очередного HTTPS запроса осуществляется из обработчика результатов предыдущего для обеспечения правильной последовательности действий, а в случае использования Websocket заранее создается набор обработчиков для всех используемых типов сообщений.

3. Реализация платформы как облачного сервиса

Платформа разработки СОЗ, как веб-ориентированный программный комплекс, может быть опубликована в рамках инфраструктуры Яндекс.Облака несколькими способами:

1. Виртуальная машина — данный вариант не имеет существенных преимуществ перед обычным «хостингом», не считая наличия услуг по управлению и обслуживанию виртуальных машин, БД и компоненты платформы размещаются на одной виртуальной машине.
2. Группа виртуальных машин — данный вариант обеспечивает возможность обработки изменения входящей нагрузки, количество машин в группе может автоматически увеличиваться при ее увеличении и уменьшаться при ее снижении. Нет возможности разместить БД на одной виртуальной машине с компонентами.

3. «Бессерверная» (*serverless*) реализация на основе «облачной функции» (*cloud function* [9]) — данный вариант характеризуется готовой средой исполнения, и, поэтому, требует меньших усилий при подготовке к работе и обслуживанию в сравнении с вариантами 1 и 2. Есть возможность автоматической обработки увеличения нагрузки — количество экземпляров функции может быть увеличено до 10, а запущенный экземпляр не останавливается сразу, а некоторое время остается доступным для следующего обращения. Нет возможности разместить БД вместе с компонентами, нет доступа к внутренней сети облака, можно использовать только фиксированный набор языков программирования, причем только один для каждой из функций.

4. «Бессерверная» реализация на основе *API Gateway* [10] — данный вариант характеризуется возможностью декларативно описать интерфейс программного комплекса в формате Open API 3.0 (Рисунок 2), а для непосредственной реализации функций использовать как перечисленные выше варианты, файлы, размещенные в «*Object Storage*», включая Docker контейнеры в «бессерверном». Существует возможность автоматической обработки увеличения нагрузки (если таковая имеется у выбранного варианта реализации). Нет возможности разместить БД вместе с компонентами, нет доступа к внутренней сети облака. Есть возможность использования различных языков программирования для реализации функций платформы.

```

openapi: 3.0.0
info:
  title: KBS Platform
  version: 1.0.0
servers:
  - url: https://d5dv0rhg35775jt1vh8m.apigw.yandexcloud.net
paths:
  /RetrieveInitializationParameters:
    get:
      x-yc-apigateway-integration:
        type: cloud_functions
        function_id: d4e21p3tolc0h0mvjipf
        tag: $latest
        service_account_id: ajeesfqv5m33t3ap6811
        operationId: kbs-platform-retrieve-ui-parameters-get
    post:
      x-yc-apigateway-integration:
        type: cloud_functions
        function_id: d4e21p3tolc0h0mvjipf
        tag: $latest
        service_account_id: ajeesfqv5m33t3ap6811
        operationId: kbs-platform-retrieve-ui-parameters-post
  /RetrieveData:
    get:
      x-yc-apigateway-integration:
        type: cloud_functions
        function_id: d4e21p3tolc0h0mvjipf
        tag: $latest
        service_account_id: ajeesfqv5m33t3ap6811
        operationId: kbs-platform-retrieve-ui-parameters-get

paths:
  /{file+}:
    get:
      x-yc-apigateway-integration:
        operationId: common-ui-components
        parameters:
          - explode: false
            in: path
            name: file
            required: true
            schema:
              type: string
            style: simple
        components:
          securitySchemes:
            httpBearerAuth:
              x-yc-apigateway-authorizer:
                type: function
                function_id: d4empde63oljiugnfv09
                authorizer_result_ttl_in_seconds: 7200
              scheme: bearer
              type: http
          x-yc-apigateway:
            service_account_id: ajehcaavm4u23qqor6lj
  
```

Рисунок 3: Фрагменты спецификации *API Gateway*, слева реализация функциональности на основе «облачной функции», справа использование «*Object Storage*» для хранения элементов UI и настройки авторизации

Для данной работе был выбран вариант «бессерверной» реализация на основе *API Gateway*, с использованием «облачных функций» для реализации функциональности. Хранение данных организовано на базе кластера PostgreSQL (*Managed Service for PostgreSQL*), по аналогии с изначальной реализацией программного комплекса. В качестве преимуществ выбранного варианта реализации можно выделить декларативную спецификацию, возможность использования возможностей *API Gateway* для реализации рутинных, но важных функций, например авторизацию, возможность использования протокола Websocket и «очереди сообщений» (Message Queue) для организации обмена данными. Кроме того, в рамках *API Gateway* предоставляется служебное доменное имя, которое может быть использовано при проведении тестов.

Возможность использования компонентов платформы в качестве отдельных веб-приложений, а также специфика «облачных функций», которые используют среду PHP,

обуславливают необходимость преобразования входных данных «облачной функции» к виду, используемому в компоненте. Для решения этой проблемы предложено использование специализированного компонента-оболочки, который обеспечивает передачу полученных данных в компонент, а также преобразование результатов работы компонента в формат выходных данных «облачной функции». В общем случае предложенный компонент-оболочка играет роль шаблона, на основе которого осуществляется создание «облачной функции» на основе компонента.

4. Заключение

В данной работе обсуждаются задачи, реализации облачного сервиса на основе программной платформы для разработки СОЗ, а также создания пользовательского интерфейса.

На платформе обеспечена возможность использования пользовательского интерфейса, предоставляемого компонентом автоматизированного формирования пользовательского интерфейса. Данный компонент реализует интерфейс во всех функциях приложения, начиная от перечня его основных функций до диалога с пользователем для ввода недостающих данных, например, по запросу от выполняемой композитной операции. Подобный подход дает возможность, во-первых, закрыть трудоемкую задачу создания пользовательского интерфейса СОЗ, а во-вторых, обеспечить возможность управления поведением клиентской части СОЗ за счет выполнения Javascript-кода, сгенерированного на основе композитной операции.

Использование предложенного компонента-оболочки, как шаблона для реализации компонентов платформы, обеспечивает упрощение процесса разработки, за счет возможности быстрого создания прототипа и применения возможностей уже существующих компонентов платформы. При этом после реализации требуемой функциональности и тестирования есть возможность оформления ее в виде облачной функции или «бессерверного» контейнера для экономии ресурсов.

Реализация платформы для разработки СОЗ в виде облачного сервиса позволяет одной стороне реализовать доступ к функциональности программного комплекса по схеме platform as a service с минимальными затратами на поддержку инфраструктуры, включая вопросы безопасности и обработки меняющейся нагрузки, а с другой при необходимости обеспечить возможность отчуждения разработанной СОЗ путем автоматизированного создания копии в заданном пользователем облаке.

5. Благодарности

Результаты получены в рамках госзадания Минобрнауки России по проекту «Методы и технологии облачной сервис-ориентированной цифровой платформы сбора, хранения и обработки больших объемов разноформатных междисциплинарных данных и знаний, основанные на применении искусственного интеллекта, модельно-управляемого подхода и машинного обучения» (№ гос регистрации: 121030500071-2).

6. Литература

- [1] HiAsm – среда разработки приложений, ориентированная на графическое программирование, <https://hiasm.com/>
- [2] LabVIEW – это среда разработки и платформа для выполнения программ, созданных на графическом языке программирования «G», <https://www.ni.com/en/shop/labview.html>
- [3] Nikolaychuk O.A., Pavlov A.I., Stolbov A.B.: The software platform architecture for the component-oriented development of knowledge-based systems. In: Proceedings of the 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1234–1239. IEEE (2018). <https://doi.org/10.23919/MIPRO.2018.8400194>.

- [4] Павлов А.И., Николайчук О.А., Столбов А.Б. Редактор концептуальных моделей: Свидетельство о государственной регистрации программ для ЭВМ №2016617626 от 11.07.2016 М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2016.
- [5] Pavlov A.I., Stolbov A.B. Domain-oriented specialization tools for knowledge-based systems development platform // CEUR Workshop Proceedings: 3rd Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS 2020; Irkutsk, 3 September 2020). 2020. Vol. 2677.
- [6] Павлов А.И., Николайчук О.А., Столбов А.Б. Web-ориентированный редактор производственных правил: Свидетельство о государственной регистрации программ для ЭВМ №2016663618 от 13.12.2016 М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2016.
- [7] Pavlov A.I., Stolbov A.B., Dorofeev A.S. The workflow component of the knowledge-based systems development platform // Proc. of 2nd Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS'2019). 2019. Vol. 2463. pp. 47-58.
- [8] Cross-platform VueJs framework, <https://quasar.dev>
- [9] Сервис «Cloud Functions», <https://yandex.cloud/ru/docs/functions/concepts/function>
- [10] Сервис для управления API-шлюзами, <https://yandex.cloud/ru/docs/api-gateway/>

Метод композиции сервисов обработки пространственных данных

Роман Константинович Фёдоров ¹

¹ *Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия*

Аннотация

Разработаны оригинальные эвристические алгоритмы и метод создания композиций сервисов, который в отличие от существующих методов проводит комплексный анализ метаданных, онтологий, экспертных знаний и статистики применения сервисов, что позволяет находить композиции сервисов на основе комбинации данных. Использование метода может значительно упростить работу пользователя и автоматизирует часто повторяющиеся его действия. В результате применения метода создания композиций опыт применения пользователем сервисов автоматически распространяется среди всех пользователей.

Ключевые слова

SOA, OGC, WPS, обработка пространственных данных, композиция сервисов

1. Введение

Для современного этапа решения крупномасштабных фундаментальных и прикладных задач актуально повышение эффективности и надежности процессов обработки и передачи данных в вычислительных машинах, комплексах и компьютерных сетях путём внедрения сервис-ориентированной архитектуры (СОА, англ. service-oriented architecture, SOA), на основе которой реализуют интеллектуальные технологии и машинное обучение, системы обработки больших объемов пространственных данных (ПД). Активно растет количество сервисов, реализующих предоставление, обработку и публикацию данных. Например, это сервисы предоставления данных дистанционного зондирования земли, обработки пространственных данных, расшифровки генома и т. д. Созданные сервисы значительно упрощают решение многих задач. В области обработки ПД определены и активно используются стандарты Open Geospatial Consortium (OGC). При этом возникает ряд сложных научно-технических задач нахождения сервисов, построения их связей между собой, проверки корректности совместного выполнения этих сервисов и др.

Объединение этих сервисов, т. е. создание композиции сервисов, позволяет решать большое количество задач. СОА значительно упрощает и ускоряет интеграцию программного обеспечения, созданного разработчиками из разных предметных областей, за счет упрощения и стандартизации интерфейсов. Композиции сервисов обеспечивают повышение уровня автоматизации решения задач, начиная от ввода данных и заканчивая публикацией результатов.

В то же время создание композиций сервисов является нетривиальной задачей. Большое количество сервисов, с одной стороны, увеличивает возможности исследователей, а с другой стороны, значительно усложняет поиск нужных сервисов для решения конкретной задачи. Составление найденных сервисов в композиции в некоторых случаях может оказаться комбинаторно сложной задачей, при этом возможна генерация достаточно большого числа

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: fedorov@icc.ru (A. 1)

ORCID: 0000-0002-2944-7522 (A. 1)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.22

альтернативных композиций сервисов. Все эти альтернативы необходимо будет оценить, выбрать наиболее релевантные для решаемой задачи, а также их проверить на возможность взаимодействия. Часто композиция из двух потенциально возможных сервисов не реализуема из-за того, что структуры данных и форматы входных и выходных данных отличаются.

Композиции сервисов позволяют объединить результаты работы исследователей, но создание таких композиций в силу сложного процесса требует высокой квалификации. Поэтому актуальной научной проблемой является разработка нового метода автоматизации построения и применения композиций сервисов.

2. Обзор методов композиции сервисов

Одна из важнейших особенностей создания композиции сервисов – это спецификация того, как результаты конкретного сервиса могут быть приняты входами других сервисов, т. е. сопоставление данных. Композиция сервисов обычно выполняется как обратный вывод, т. е. от целевого состояния к исходному состоянию. Существует несколько подходов к автоматизации создания композиций сервисов [1].

1. Интерактивное построение композиций сервисов на основе знаний. В этом подходе пользователь вручную строит композицию сервисов, а система вывода и знания (метаописания, онтологии) используется для поиска подходящих сервисов и значений параметров [2, 3, 4].

2. Автоматическое построение композиций сервисов. В этом подходе производится (полу) автоматическое обнаружение и составление необходимых сервисов на основе семантического анализа и логического вывода [5, 6, 7].

3. Автоматическая композиция Web-сервисов с учетом планирования выполнения. В этом подходе [8, 9] входные данные представляются сервисами, которые семантически аннотированы с использованием онтологий, например Web Ontology Language for Services (OWL-S) [10], сервисы обработки аннотированы с помощью Web Service Modeling Ontology (WSMO) [11]. Выбор среди сгенерированных цепочек сервисов может проводиться на основе оценки качества обслуживания (QoS) с использованием алгоритмов, таких как генетические алгоритмы или алгоритмы теории игр [12, 13].

В последние годы многие исследования подчеркивали важность семантической информации о сервисе при композиции сервисов геообработки, утверждая, что при составлении сервисов необходимо обнаружение сервисов с соответствующими функциями [14, 15, 16]. Входы и выходы между соседними сервисами должны совпадать не только с точки зрения параметров, но также с точки зрения их синтаксиса и семантики. Онтология стала распространенным инструментом для описания семантической информации сервисов геообработки, а онтологические подходы в настоящее время являются наиболее распространенными методами автоматической компоновки сервисов геообработки.

Следует отметить, что автоматическое построение композиций сервисов на основе семантического анализа, логического вывода и планирования является сложной задачей по следующим причинам:

- большая часть сервисов имеет недостаточно метаинформации, либо она отсутствует или не формализована;
- требуется проработанная онтология предметной области, ее построение не тривиальный процесс, в котором нужно рассмотреть все возможные связи между сервисами. Достаточно сложно выполнить описание некоторых сервисов. В статье [6] приводится пример, в котором сервис, выполняющий деление, может возвращать плотность, если его входные данные представляют собой массу и объем, или скорость, если задано расстояние и период времени;
- требуется согласованность метаинформации сервисов и онтологии предметной области.

Поиск и создание новых композиций сервисов является сложной задачей, в которой необходимо проанализировать большое количество сервисов и предоставить наиболее релевантные их комбинации. Существующие методы и подходы на основе анализа метаданных и онтологий не позволяют значительно упростить поиск из-за частого отсутствия метаданных и трудоемкости разработки онтологий. Создание композиций сервисов ограничено сложностью

взаимодействия сервисов по данным из-за различий в программном интерфейсе доступа к данным, в структуре данных, используемых справочников и т. д.

2.1. Предлагаемый метод композиции сервисов

Для создания композиций сервисов требуется метод, который учитывает перечисленные требования и позволяет связывать различные сервисы между собой для достижения целей пользователей. При этом метод должен учитывать непроработанность метаданных сервисов, недостаточную онтологическую формализацию данных и процессов и давать возможность разрабатывать новые сервисы или средства обработки данных для создания композиций сервисов.

Работа метода основывается на новой вычислительной модели композиции сервисов, определяемой структурой

$CM = \langle Z, S, A, T, CR, F \rangle$ со следующими обозначениями.

Z – множество параметров. В стандарте WPS OGC обработки пространственных данных параметры связаны с определенными типами данных.

S – множество сервисов. В работе рассматриваются только сервисы без хранения состояния (stateless), т. е. результат выполнения сервисов не должен зависеть от предыдущих вызовов. Этого достаточно для решения широкого круга задач. Сервис $s \in S$ – это упорядоченная тройка $s = \langle name, Z^{Inp}, Z^{Out} \rangle$, где $name$ – это имя сервиса, $Z^{Inp} \in Z$ – множество входных параметров, $Z^{Out} \in Z$ – множество выходных параметров сервиса. Среди выходных параметров сервиса имеется параметр успешности выполнения сервиса. Далее будем обозначать $s.Z^{Inp}$ и $s.Z^{Out}$ параметры, принадлежащие определенному сервису. Сервисы разделяются на сервисы получения данных $S^d \subset S$, вычислительные сервисы $S^c \subset S$ и сервисы публикации данных $S^p \subset S$.

A – множество пользователей, которые взаимодействуют с СОИАС через применение сервисов S . Набор применяемых ими сервисов может варьироваться.

T – множество заданий выполнения сервисов из S . Каждое задание $t \in T$ определяется упорядоченной четверкой $t = \langle a, s, V_i, V_o \rangle$, т. е. вызов сервиса $s \in S$ пользователем $a \in A$ с множеством значений V_i входных параметров $s.I$ и результаты V_o выходных параметров $s.O$.

CR – множество вычислительных узлов, на которых выполняются сервисы. Каждый сервис может быть активирован на заранее определенных узлах. Автоматическое выделение новых узлов для сервисов в работе не рассматривается.

F – множество функций ранжирования сервисов.

Отметим, что композиция сервисов является также сервисом с заданным именем, входными и выходными параметрами. Общеизвестным стандартом определения композиций является использование направленного ациклического графа (Directed Acyclic Graph, (DAG)) $s_i = DAG_i = \langle T_i, E_i \rangle$, $s_i \in S$, в котором вершинами являются задания $t \in T_i$, а ребра $e = \langle v_k, v_l \rangle$, $e \in E_i$ показывают передачу данных от выходного параметра одного задания к входному параметру другого задания, где $v_k \in V_k^i$, $v_l \in V_l^j$, $t_i = \langle a_i, s_i, V_i^i, V_o^i \rangle$, $t_j = \langle a_j, s_j, V_j^j, V_o^j \rangle$. Поскольку возможна передача данных одновременно по нескольким параметрам, то ребер между двумя вершинами (заданиями) также может быть несколько. В литературе чаще всего передача данных между заданиями обычно обозначается одним ребром. Для создания готовой к выполнению композиции необходимо определить все передаваемые данные между параметрами.

Определим типы данных, используемые для обработки пространственных данных. Частично примем обозначения, принятые в реляционной алгебре:

$D = \{D_1, \dots, D_l\}$ – множество скалярных типов, определенных в стандарте WPS;

$R = \{R_1, \dots, R_m\}$ – множество схем отношений;

$d = \{d_1, \dots, d_k\}$ – множество таблиц (отношений). В качестве параметров могут использоваться реляционные таблицы $d_i \in Z$.

Введем множество типов параметров $TZ = D \cup R$. Для каждого параметра определен тип $tz_j = type(z_j)$, $tz_j \in TZ$.

Для формирования композиций сервисов требуется информация о возможности передачи данных выходного параметра сервиса для использования во входном параметре другого сервиса. Однако соответствие типов параметров не является достаточным, например, тип «число» может иметь смысл – высота над уровнем моря, а в другом случае – температура. Поэтому соответствие типов является необходимым условием, но не достаточным. Получение информации о возможности передачи данных между параметрами может производиться на основе анализа метаданных, онтологий, экспертных знаний и статистики применения сервисов и т. д. Поэтому вводим понятие метки, которое позволяет абстрагироваться от способа получения информации о возможной передаче данных между параметрами. Метка используется для идентификации возможности передачи данных, т. е. если выходной параметр и входной параметр имеет одну и ту же метку, то между ними возможна передача данных. Каждому параметру может соответствовать ноль и более меток. Метками могут являться онтологические концепты, заданные с помощью URI, или некоторые сгенерированные уникальные значения.

$m \in M$ – множество меток. Для каждого параметра z_j определено подмножество меток M_j . Проверка возможности передачи данных между сервисами в рамках модели выполняется на основе соответствия типов параметров и меток. Для двух сервисов s^i и s^j возможна передача данных от параметра $z^i \in s^i.Z^{Out}$ к параметру $z^j \in s^j.Z^{Inp}$, если типы параметров совпадают $type(z_i) = type(z_j)$ и существует метка $m \in M$, которая имеется у обоих параметров ($m \in M_i$ и $m \in M_j$).

Выполнение этого условия не гарантирует корректное выполнение сервиса s^j . Оно позволяет значительно сузить количество возможных композиций сервисов.

2.1.1. Основные этапы метода

Процесс создания композиции сервисов представлен на рисунке 2.1 и состоит из трех основных этапов.

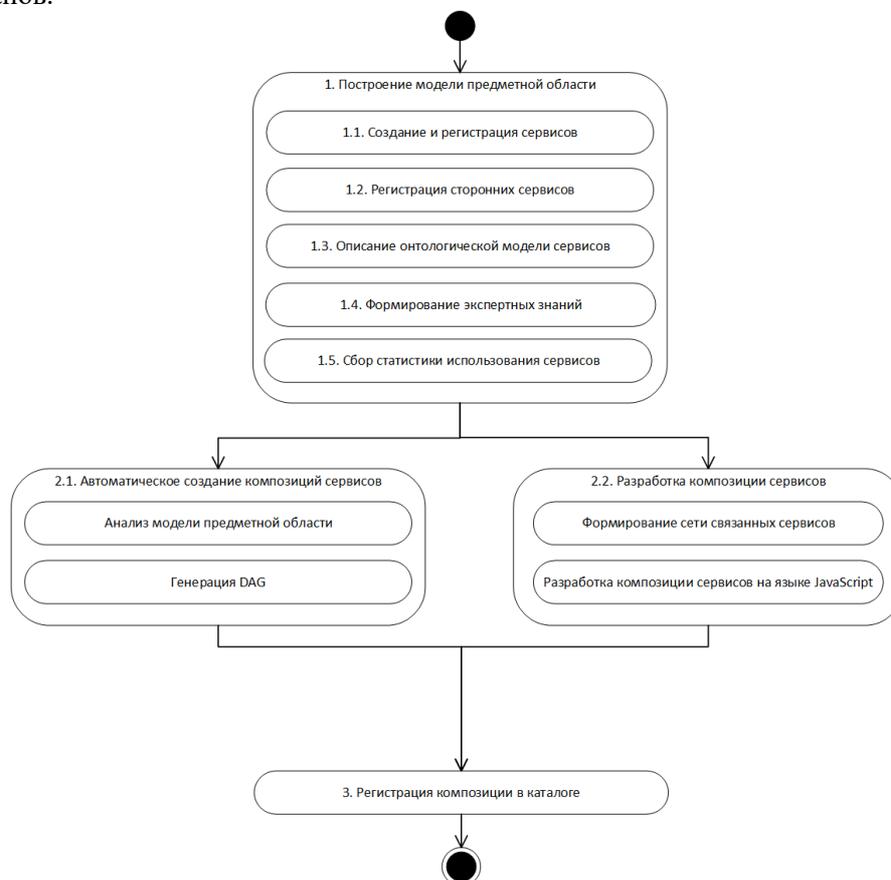


Рисунок 1: Схема метода формирования композиций сервисов

Этап 1. Построение модели предметной области

Состоит из 5 шагов, выполняемых независимо друг от друга.

1.1. Создание и регистрация сервисов – предполагается разработка вычислительных сервисов, сервисов данных и сервисов публикации данных. Вычислительные сервисы разрабатываются в соответствии со стандартом Web Processing Service (WPS, разработан Open Geospatial Consortium) на вычислительных ресурсах среды или на любых других серверах, к которым имеется сетевой доступ. Для создания сервисов данных используется компонент «Фабрика сервисов ввода и редактирования реляционных данных», который поддерживает все основные этапы работы с табличными данными. «Фабрика сервисов отображения графиков» и «Фабрика сервисов отображения пространственных данных» предназначены для создания сервисов публикации данных. Все создаваемые сервисы регистрируются в каталоге, обладают программным интерфейсом и готовы стать частью композиции сервисов.

1.2. Регистрация сторонних сервисов – осуществляется подключение WPS сервисов, созданных на сторонних вычислительных ресурсах. В каталоге сервисов указывается необходимая для запуска сервиса информация и метаданные.

1.3. Описание онтологической модели сервисов – предназначено для определения возможности передачи данных между параметрами сервисов на основе онтологий. В рамках описания онтологической модели сервисов предполагается определение концептов для параметров сервисов. Каждому параметру может соответствовать ноль и более меток. Метками могут являться онтологические концепты, заданные с помощью URI. Использование одного и того же концепта для выходного параметра сервиса и для входного параметра другого сервиса предполагает возможность передачи данных между ними и присвоение одной метки для них.

1.4. Формирование экспертных знаний – используется для определения возможности передачи данных пользователем. Разработка онтологий – достаточно сложный процесс. Часто с помощью онтологий не все сущности и отношения предметной области формализованы. Поэтому в качестве меток параметров сервисов могут использоваться некоторые сгенерированные уникальные значения, назначаемые пользователем для конкретных сервисов. Например, сервис построения карты плотности может применяться для всех точечных данных. Соответственно, выходным параметрам, которые предоставляют векторный точечный слой, можно назначить ту же самую метку, что и у входного параметра сервиса построения карты плотности.

1.5. Сбор статистики выполнения сервисов – производится сбор информации о произведенных вызовах сервисов, которые включают название сервиса и его адрес, значения входных и выходных параметров, время выполнения сервисов, успешность выполнения, ошибки выполнения и т. д. Данные сохраняются в виде таблицы 1, где значения параметров хранятся в формате JSON.

Таблица 1

Данные о вызовах сервисов

Id задания	Id сервиса	Статус	Входные параметры	Выходные параметры	Начало	Конец
3986	309	Success	{"table":"ticki"}	{"file":"http://84	18:20	18:22
3989	307	Failed	{"disp":"20"}		18:21	
3985	308	Success	{"file":"http://84..."}	{"res":"http://84	18:21	18:22
3987	307	Success	{"file":"http://84..."}	{"dist":"28",	19:18	19:19

Далее данные этой таблицы будем обозначать как Log – множество заданий (успешных вызовов сервисов), совершенных пользователями. Причиной неудачного вызова сервиса могут быть некорректные значения параметров, поэтому данные фильтруются и оставляются только успешные вызовы.

Этап 2. Создание композиций сервисов

После построения модели предметной области на этапе 2 применяются два альтернативных способа создания композиций:

- автоматическое формирование композиций сервисов. Данный способ позволяет создать композицию сервисов на основе анализа графа связности заданий;
- разработка композиций сервисов. Пользователь разрабатывает композиции сервисов с помощью языка JavaScript. Этот способ требуется при необходимости обработки промежуточных данных с помощью средств языка, или при заранее неизвестном количестве заданий в композиции DAG_i, которые формируются в зависимости от данных.

Этап 3. Регистрация композиции сервисов в каталоге

Последним этапом является сохранение полученных композиций сервисов в каталоге для последующего применения пользователями.

2.1.2. Автоматическое формирование композиций сервисов

Создание композиций сервисов должно проводиться на основе анализа возможных связей между сервисами. Применение меток обобщает использование информации о применении сервисов, экспертных знаний и онтологий. Каждая композиция задается графом DAG, с заданными параметрами заданий и связями по данным. Создаваемые композиции сохраняются в каталоге сервисов, где пользователь может найти нужную композицию в соответствии со своим запросом. Автоматическое создание композиций состоит из следующих шагов:

Шаг 1. Присвоение меток параметрам сервисов.

Шаг 2. Установление связей между сервисами и создание графа сервисов (граф DAG_{gen}).

Шаг 3. Поиск компонент связности графа DAG_{gen}.

Шаг 4. Декомпозиция графа сервисов.

2.1.3. Присвоение меток параметрам сервисов

Присвоение меток параметрам сервисов производится следующими способами:

1. Задаются вручную (экспертные знания).
2. Анализ статистики выполнения сервисов.
3. Анализ метаданных и онтологических понятий.

Рассмотрим подробнее все способы.

Задание вручную меток сервисам производится в основном для общеиспользуемых сервисов, например для всех сервисов редактирования реляционных данных ставится общая метка, которая связывает с сервисами экспорта в формат XLS, CSV. Для сервисов редактирования реляционных данных с пространственными атрибутами задается общая метка с сервисами экспорта в форматы shape, geojson и т.д.

Анализ статистики выполнения сервисов производится для того, чтобы определить факт передачи данных между параметрами сервисов. Определение связи между двумя вызовами сервисов может производиться на основе анализа значений параметров. Множество параметров можно разделить на входные и выходные. Предполагается, если значение выходного параметра совпадает со значением входного параметра, то между вызовами сервисов может быть передача данных. Значения параметров могут быть строковыми или числовыми, а также идентификаторами ресурсов URI, обычно это URL ссылки на файлы, сервисы предоставления данных и т. д. Каждый URI, используемый для получения данных WPS сервисом, однозначно идентифицирует передаваемые данные. Соответственно, по URI параметрам можно однозначно определить, что данные передаются от одного вызова сервиса другому. Несколько сложнее со строковыми и числовыми параметрами, совпадение значений которых может быть случайным. На текущий момент строковые и числовые параметры не рассматриваются. Если факт передачи данных между параметрами сервисов определен, то генерируется уникальная метка, которая добавляется указанным параметрам.

Анализ метаданных и онтологических понятий проводится пока только для сервисов, предоставляющих реляционные данные. Между данными часто имеется отношение общее – частное. В терминах объектно-ориентированного программирования это отношение называется наследованием (суперкласс – подкласс). Например, собираемые данные о конференциях, семинарах, лекциях обобщаются данными, описывающие классом мероприятие (название,

место, дата, организатор и т.д.). Применение какой-либо обработки к данным суперкласса за редким исключением, означает возможность ее применения к данным подклассов. Например, если обработка производит подсчет и строит график распределения мероприятий по годам, то ее можно применить к данным, описывающим конференции при условии соблюдения схемы данных суперкласса. Создан рекурсивный алгоритм, который все метки общего класса распространяет всем дочерним классам.

Отношение между данными суперкласс – подкласс определяется разными способами:

1. Получаются в результате создания сервисов данных на основе других.
2. Проводится анализ онтологических понятий метаданных сервисов.

Для описания понятий используется язык OWL (Web Ontology Language) [17], разработанный консорциумом W3C. Язык OWL позволяет описывать классы и отношения между ними. В основе языка – представление действительности в модели данных «объект – свойство». Язык OWL позволяет определить иерархию классов, включая суперклассы, под которыми объединяются подклассы, что позволяет применять правило ко всем классам. Определение иерархии классов задается с помощью элементов `subClassOf`. При анализе метаданных при наличии онтологических понятий производится запрос на наличие отношения `subClassOf` между двумя классами, если оно имеется, то применяется алгоритм распространения меток.

2.1.4. Установление связей между сервисами и создание графа сервисов

В графе DAG_{gen} вершинами являются сервисы, зарегистрированные в каталоге. Дуги графа являются направленными и определяют связь (возможность передачи данных) между параметрами сервисов. Для поиска дуг производится поиск одинаковых меток у параметров сервисов. Если такие пары параметров найдены, то добавляется дуга в граф DAG_{gen} .

2.1.5. Поиск компонент связности

Граф сервисов DAG_{gen} содержит возможные применения для решения множества разных задач. Следовательно, его необходимо декомпозировать и выделить отдельные подграфы, соответствующие решаемым задачам, т.е. композиция сервисов образуют связанный подграф D_i (компонента связности графа) в DAG_{gen} . Для получения подграфа, соответствующего решению задачи, недостаточно найти компоненты связности в DAG_{gen} , так как в нем могут быть сервисы получения данных. Например, могут быть сервисы получения данных ДЗЗ или сервисами предоставления векторных и реляционных данных. Эти сервисы могут использоваться для решения большого количества различных задач. Например, данные ДЗЗ Landsat 8, которые можно получить с помощью сервисов, используются для задач оценки подтоплений из-за разлива рек, динамики изменения ледников, классификации породного состава лесов и т. д. Каждому решению задачи будет соответствовать свой подграф в DAG_{gen} , и они будут все связаны между собой через сервисы получения данных ДЗЗ, т. е. объединяться в компоненты связности. Поэтому при выделении подграфов, соответствующих только одной задаче, необходимо не рассматривать связи через сервисы получения данных.

Для поиска связанных подграфов D_i применяется алгоритм поиска в ширину. При этом сложность поиска является линейной от суммы числа вершин и числа рёбер графа DAG_{gen} . Это обосновывается направленностью графа и отсутствием циклов.

2.1.6. Декомпозиция графа сервисов

Алгоритм поиска компонент связности может получить достаточно много подграфов D_i . Часть подграфов может соответствовать экспериментам, которые привели к недостаточным результатам. Также среди них может быть достаточно много дублированных, которые возникают из-за многократного решения задачи на разных данных и с разными параметрами. Поэтому производится фильтрация множества D_i и выделение наиболее полезных композиций сервисов.

3. Апробация

Данные гербарных листов определены стандартами TDWG Darwin Core и Access to Biological Collections Data (ABCD). На основе этих стандартов создан сервис сбора и редактирования реляционных данных «BasicHerbarium». «BasicHerbarium» стал базовым классом для следующих сервисов: БД «Гербарий сосудистых растений Сибирского института физиологии и биохимии растений СО РАН» (IRK); БД «Гербарий мхов Сибирского института физиологии и биохимии растений СО РАН» (IRK) и т.д.

Пользователь для получения карты распространения видов последовательно использовал сервисы:

1. «BasicHerbarium».
2. table2shp – сохранил данные сервиса «BasicHerbarium» в формат SHAPE.
3. vector2grid52 – произвел обработку данных в формате SHAPE для получения растрового файла в формате GeoTIFF. В каждой ячейке (пикселе) записано количество точек.
4. map – создал карту распространения видов с указанием легенды (задание цветовой палитры отображений файла GeoTIFF в зависимости количества точек в ячейке).

На основе анализа статистики создана композиция сервисов densityMap (рисунок 2).

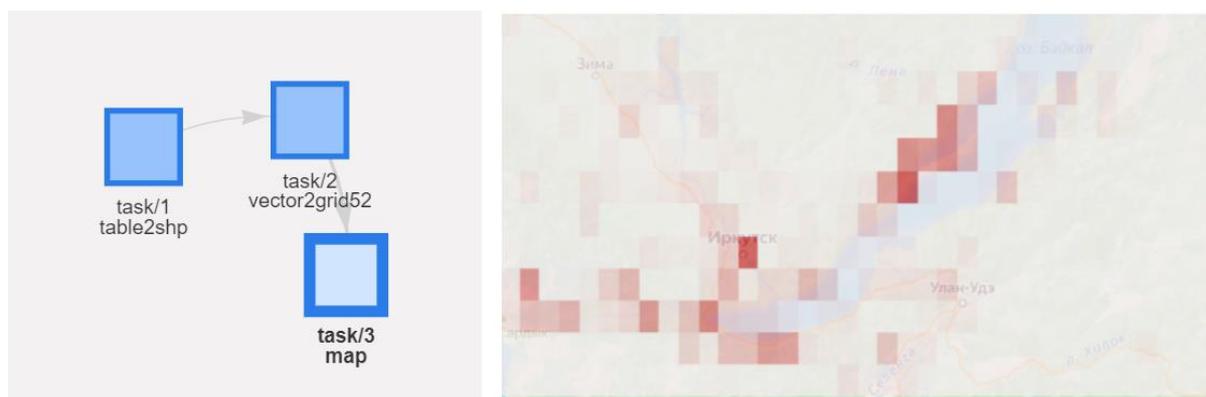


Рисунок 2: Сервис densityMap и результат его работы

Созданный сервис автоматически предлагается всем пользователям сервисов данных, созданных на основе «BasicHerbarium». При выборе этого сервиса автоматически создается новая композиция сервисов002E

4. Заключение

Разработаны оригинальные эвристические алгоритмы и метод создания композиций сервисов, базирующийся на применении предложенной модели, который в отличие от существующих методов проводит комплексный анализ метаданных, онтологий, экспертных знаний и статистики применения сервисов, что позволяет находить композиции сервисов на основе комбинации данных. Использование метода может значительно упростить работу пользователя и автоматизирует часто повторяющиеся его действия. В результате применения модели и метода создания композиций опыт применения пользователем сервисов автоматически распространяется среди всех пользователей.

Работа выполнена при поддержке ЦКП ИИВС ИРНОК, базового проекта 121030500071-2.

5. Список литературы

- [1] Zhi-Wei H., Cheng-Zhi Q., A-Xing Z., Peng L., Yi-Jie W., Yun-Qiang Z. From Manual to Intelligent: A Review of Input Data Preparation Methods for Geographic Modeling // ISPRS

- International journal of geo-information. 2019. Vol. 8, № 9. article 376. DOI: 10.3390/ijgi8090376.
- [2] Di L., Zhao P., Yang W., Yue P. Ontology-Driven Automatic Geospatial-Processing Modeling Based on Web-Service Chaining // Proceedings of the Sixth Annual NASA Earth Science Technology Conference. College Park, MD, USA. 2006. P. 27-29.
 - [3] Zhao P., Di L., Yu G., Yue P., Wei Y., Yang W. Semantic Web-based geospatial knowledge transformation // Computers & Geosciences. 2009. Vol. 35, № 4. P. 798-808.
 - [4] Scheider S., Ballatore A. Semantic typing of linked geoprocessing workflows // International Journal of Digital Earth. 2017. Vol. 11. P. 113-138.
 - [5] Jiang J., Zhu A.X., Qin C.Z., Zhu T., Liu J., Du F., Liu J., Zhang G., An Y. CyberSoLIM: A cyber platform for digital soil mapping // Geoderma. 2016. № 263. P. 234–243.
 - [6] Lutz M., Lucchi, R., Friis-Christensen A., Ostländer N. A Rule-Based Description Framework for the Composition of Geographic Information Services // Proceedings of the International Conference on GeoSpatial Semantics. Mexico City, Mexico, 29–30 November. 2007. P. 114-127.
 - [7] Lutz M. Ontology-based descriptions for semantic discovery and composition of geoprocessing services // GeoInformatica. 2007. Vol. 11. P. 1-36.
 - [8] Yue P., Di L., Yang W., Yu G., Zhao P., Gong J. Semantic Web Services-based process planning for earth science applications // International Journal of Geographical Information Science. 2009. Vol. 23. P. 1139-1163.
 - [9] Farnaghi M., Mansourian A. Automatic composition of WSMO based geospatial semantic web services using artificial intelligence planning // Journal of Spatial Science. 2013. Vol. 58. P. 235-250.
 - [10] Martin D., Burstein M., Hobbs J., Lassila O., McDermott D., McIlraith S., Narayanan S., Paolucci M., Parsia B., Payne T. OWL-S: Semantic markup for web services // W3C Member Submission. 2004.
 - [11] Roman D., Keller U., Lausen H., Bruijn J.D., Stollberg M., Polleres A., Feier C., Bussler C., Fensel D. Web Service Modeling Ontology // Applied ontology. 2005. № 1. P. 77-106.
 - [12] Li H., Zhu Q., Yang X., Xu L. Geo-information processing service composition for concurrent tasks: A QoS-aware game theory approach // Computers & Geosciences. 2012. Vol. 47. P.46-59.
 - [13] Yue P., Tan Z., Zhang M. GeoQoS: Delivering Quality of Services on the Geoprocessing Web // Proceedings of the OSGeo's European Conference on Free and Open Source Software for Geospatial (FOSS4G-Europe 2014). Bremen, Germany. 15–17 July 2014.
 - [14] Cruz S.A., Monteiro A.M., Santos R. Automated geospatial web services composition based on geodata quality requirements // Computers & Geosciences. 2012. Vol. 47. P. 60-74.
 - [15] Huang W., Harrie L. Towards knowledge-based geovisualisation using Semantic Web technologies: A knowledge representation approach coupling ontologies and rules // International Journal of Digital Earth. 2020. Vol.13. P. 976-997.
 - [16] Sun Z., Yue P., Lu X., Zhai X., Hu L. A task ontology driven approach for live geoprocessing in a service-oriented environment // Transactions in GIS. 2012. Vol. 16. P. 867-884.
 - [17] Web Ontology Language (OWL) [Электронный ресурс]. – Режим доступа: <https://www.w3.org/OWL/> (дата обращения 25.02.2024).

Методы и средства разработки и применения испытательного стенда сервис-ориентированных приложений

Александр Феокистов¹, Михаил Воскобойников¹, Алексей Еделев^{1,2}

¹ Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

² Институт систем энергетики им. Л.А. Мелентьева СО РАН, ул. Лермонтова, д. 130, 664033, Иркутск, Россия

Аннотация

Рассматривается подход к разработке и применению испытательного стенда сервис-ориентированных приложений в гетерогенной распределенной вычислительной среде. Целью испытательного стенда является обеспечение разработчиков средствами проведения экспериментов по оценке качества данных, работы алгоритмов, анализа результатов вычислений и других аспектов реализации сервис-ориентированных приложений. Сервис-ориентированное приложение создается на основе рабочих процессов, представляемых в виде композиции веб-сервисов. Веб-сервисы реализуют выполнение прикладного и системного программного обеспечения. Прикладное программное обеспечение поддерживает выполнение операций предметной области приложения. Суть и новизна подхода заключается в использовании в рамках испытательного стенда системного программного обеспечения, предоставляющего дополнительные функции обработки и анализа данных, получения информации с контрольно-измерительного оборудования и многокритериального выбора оптимальных результатов вычислений. Рабочие процессы могут быть представлены в стандартной форме на декларативном языке Business Process Execution Language для использования во внешних системах управления рабочими процессами.

Ключевые слова

Научные приложения, рабочие процессы, сервисы, испытательные стенды

1. Введение

В настоящее время чрезвычайную актуальность приобретает исследование современного состояния природно-технических систем (ПТС) [1]. Природно-техническая система представляет собой совокупность форм и состояний различных процессов взаимодействия компонентов природной среды с инженерными сооружениями на всех стадиях ее функционирования [2]. К таким системам относятся, например, энергетические комплексы на разных уровнях их территориально-отраслевой иерархии, включая микросети различных инфраструктурных объектов. Как правило, в рамках подобных исследований рассматривается оптимизация взаимодействия между естественными и искусственными компонентами ПТС [3].

В рамках цифрового мониторинга Байкальской природной территории (БПТ) [4] разрабатываются новые модели, алгоритмическое и программное обеспечение (ПО), научные сервисы, инструментальные средства и методики исследования функционирования и развития ПТС, располагающихся на БПТ. В частности, данная работа посвящена методам и средствам разработки и применения испытательного стенда сервис-ориентированных приложений для

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: agf@icc.ru (A. 1); mikev1988@mail.ru (A. 2); flower@isem.irk.ru (A. 3)

ORCID: 0000-0002-9127-6162 (A. 1); 0000-0003-3034-4907 (A. 2); 0000-0003-2219-9754 (A. 3)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.23

моделирования, структурно-параметрической оптимизации, экологической и технико-экономической оценки ПТС. Основной акцент делается на сервис-ориентированные приложения, создаваемые на основе рабочих процессов в виде композиции веб-сервисов.

Существует широкий спектр систем управления рабочими процессами (см., например, краткий обзор в [5]), в том числе систем, поддерживающих сервис-ориентированные рабочие процессы [6-8]. Однако средства тестирования и испытания компонентов сервис-ориентированных приложений в этих системах слабо развиты. В этой связи в рамках данного исследования предложены новые методы и средства разработки и применения испытательного стенда сервис-ориентированных приложений в гетерогенной распределенной вычислительной среде. Применение предложенных разработок проиллюстрировано на примере исследования работы модуля моделирования функционирования энергетических комплексов, относящихся к разным уровням их территориально-отраслевой иерархии.

2. Испытательный стенд

Термин *испытательный стенд* (англ., *testbed*) [9] широко используется в научных исследованиях, в частности, в рамках создания, развития и использования информационно-вычислительных и телекоммуникационных систем и технологий, для определения программно-аппаратной среды, специализированной для проведения экспериментов, связанных с тестированием и испытанием новых алгоритмических и программных разработок, различных технологических решений и источников данных, а также систем их извлечения, обработки, передачи, хранения и анализа, перед внедрением создаваемых или модифицируемых систем и технологий. Испытательные стенды создаются в физической или виртуальной формах с целью обеспечения разработчиков контролируемой средой для проверки работоспособности, удовлетворения функциональным и системным требованиям, производительности, эффективности использования ресурсов и других свойств исследуемых систем и технологий.

В рамках данного исследования испытательный стенд сервис-ориентированных приложений реализован с помощью инструментального комплекса Framework for Development and Execution of Scientific WorkFlows (FDE-SWFs) [10] в виде НПП, включающего сервисы для выполнения прикладного и системного программного обеспечения. FDE-SWFs поддерживает и существенно развивает набор функциональных возможностей программного комплекса Orlando Tools [11].

Архитектура FDE-SWFs включает следующие основные компоненты: пользовательский интерфейс, конструктор вычислительной модели разрабатываемого приложения, конструктор модулей и сервисов их выполнения, менеджер ПО, менеджер вычислений, менеджер вычислительных процессов и состояния среды, API для доступа к информационно-вычислительным ресурсам и системам.

Общая структура испытательного стенда, разрабатываемого с помощью FDE-SWFs, представлена на рисунке 1. Стенд создается в виде рабочего процесса, дополненного набором системных операций. Рабочий процесс выполняется менеджером вычислений FDE-SWFs.

Рабочий процесс представляется на стандартизированном декларативном языке Business Process Execution Language (BPEL) [12] и в дальнейшем может использоваться любыми внешними системами управления рабочими процессами, поддерживающими BPEL. BPEL определяет модель для описания поведения сервис-ориентированных приложений в терминах взаимодействий (совокупности сообщений) между внутренними вычислительными процессами и их партнерами (внешними сервисами). В рамках сервис-ориентированной архитектуры применяются сервисы, представляющие модули разрабатываемого приложения и внешние приложения. Внешние приложения реализуют дополнительные системные операции обработки и анализа данных, получения информации с контрольно-измерительного оборудования и многокритериального выбора оптимальных результатов вычислений. В целом веб-сервис представляет программную систему со стандартизированными интерфейсами, идентифицируемую уникальным URL-адресом. При этом веб-сервисы не обладают знаниями о выполняемых приложениях, а приложения не нуждаются в информации о способах их выполнения веб-сервисами.

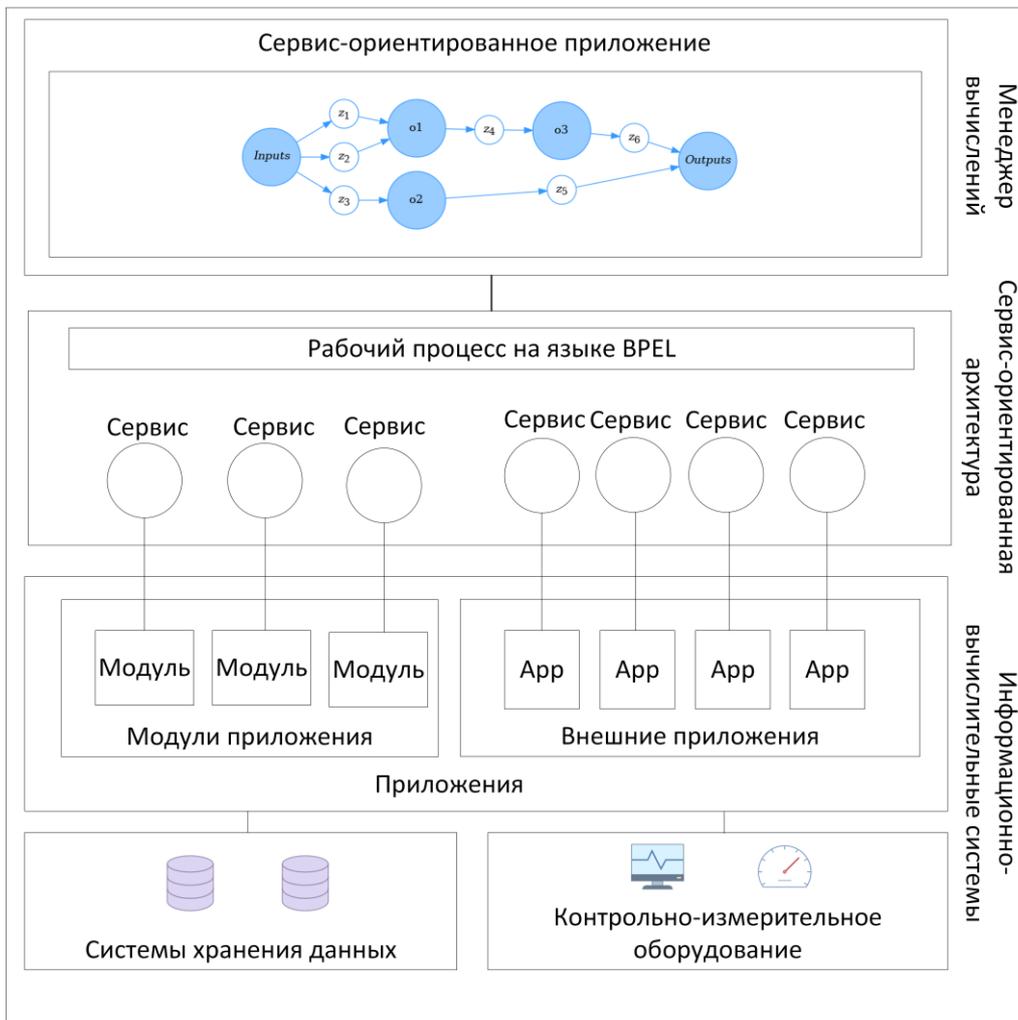


Рисунок 1: Структура испытательного стенда

3. Пример

В качестве примера применения испытательного стенда исследован набор методов структурно-параметрической оптимизации модели энергетического комплекса с помощью библиотеки `ragto` [13]. Библиотека `ragto`, реализованная на языке C++, предназначена для параллельной оптимизации. Она предоставляет унифицированный интерфейс для алгоритмов оптимизации с целью упрощения их развертывания в вычислительных средах с массовым параллелизмом.

Исследовано 60 методов библиотеки `ragto` с их настройками, заданными по умолчанию. Результаты сравнительного анализа методов структурно-параметрической оптимизации энергетического комплекса национального уровня территориально-отраслевой иерархии представлены на рисунке 2.

За допустимое время решения задачи полностью отработали пять методов: (N+1)-ES Simple Evolutionary Algorithm (m1), Augmented Lagrangian algorithm (m2), Method of Moving Asymptotes (m3), Algorithm AGS to solve constrained nonlinear programming problems with Lipschitzian functions (m4) и Interior Point OPTimizer (m5). Среднее время выполнения метода, средний размер используемой им оперативной памяти и средняя загрузка процессора были получены путем профилирования процесса выполнения модуля для конкретного метода.

На основе многокритериального выбора единственного варианта значений критериев по всем трем правилам (Парето-оптимальному, мажоритарному и лексикографическому) наиболее приемлемым методом является метод m5.

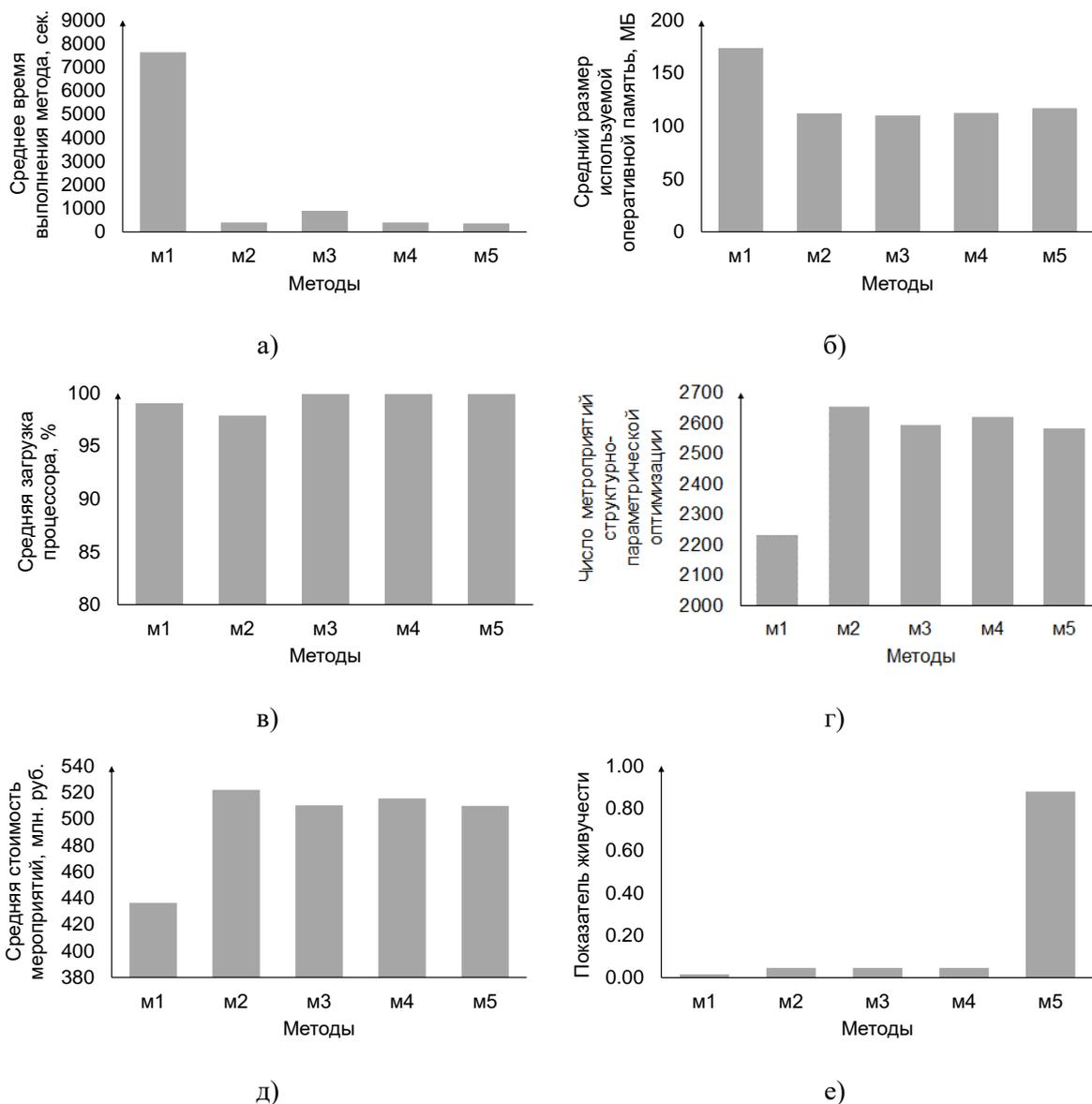


Рисунок 2: Результаты сравнительного анализа по следующим критериям: среднее время выполнения метода (а), средний размер используемой оперативной памяти (б), средняя загрузка процессора (в), число мероприятий структурно-параметрической оптимизации (г), средняя стоимость мероприятий (д), показатель живучести (е)

4. Заключение

В рамках исследования предложены новые методы и средства разработки и применения испытательного стенда сервис-ориентированных приложений. Разработан испытательный стенд для исследования работы модуля моделирования энергетических комплексов, относящихся к разным уровням их территориально-отраслевой иерархии. Модуль реализует набор методов структурно-параметрической оптимизации модели энергетического комплекса с помощью библиотеки ragto.

Стенд ориентирован на тестирование сервиса для выполнения данного модуля на вычислительных ресурсах с разными вычислительными характеристиками. Применение стенда позволило осуществить выбор алгоритмов, оптимальных по заданному набору критериев.

5. Благодарности

Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды», рег. № 124052100088-3).

6. Список литературы

- [1] С. В. Исаев, Концепция природно-технических систем и ее использование при изучении антропогенной трансформации природной среды, Географический вестник, 2016, № 3 (38), с. 105–113. doi:10.17072/2079-7877-2016-3-105-113.
- [2] А. Л. Ревзон, Картография состояния природно-технических систем, М., Недра, 1992.
- [3] Т. К. Салина, Концептуальный подход к управлению устойчивым развитием ТЭК Северо-Западного федерального округа, Журнал экономической теории, 2016, № 4, с. 80–91.
- [4] Цифровой мониторинг Байкальской природной территории. URL: <https://baikal-project.icc.ru/>.
- [5] A. Feoktistov, A. Edelev, A. Tchernykh, S. Gorsky, O. Basharina, and E. Fereferov, An Approach to Implementing High-Performance Computing for Problem Solving in Workflow-based Energy Infrastructure Resilience Studies, *Computation*, 2023, vol. 11, no. 12, pp. 243. doi: 10.3390/computation11120243.
- [6] J. Goecks, A. Nekrutenko, J. Taylor, and Galaxy Team, Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences, *Genome Biol*, 2010, vol. 11, pp. 1–13. doi:10.1186/gb-2010-11-8-r86.
- [7] V. Balis, HyperFlow: A model of computation, programming approach and enactment engine for complex distributed workflows, *Future Generation Computer Systems*, 2016, vol. 55, pp. 147–162. doi:10.1016/j.future.2015.08.015.
- [8] M. H. Hilman, M. A. Rodriguez, and R. Buyya, Workflow-as-a-service cloud platform and deployment of bioinformatics workflow applications, in: *Knowledge Management in the Development of Data-Intensive Systems*, I. Mistrik, M. Galster, B. Maxim, B. Tekinerdogan (Eds.), CRC Press, Boca Raton, Florida, US, 2021, pp. 205–226.
- [9] Г. И. Радченко, Распределенные виртуальные испытательные стенды: использование систем инженерного проектирования и анализа в распределенных вычислительных средах, Вестник Южно-Уральского государственного университета. Серия: Математическое моделирование и программирование, 2011, № 37 (254), с. 108–121.
- [10] G. Danilov and M. Voskoboinikov, Testbed-based approach to testing a library for evaluating network reliability algorithms, *Proceedings of the International Workshop on Critical Infrastructures in the Digital World (IWCI-2024)*, March 14-20, 2024, Bolshoe Goloustno, Russia, Irkutsk, ESI SB RAS, 2024, pp. 3-4.
- [11] A. Feoktistov, S. Gorsky, I. Sidorov, I. Bychkov, A. Tchernykh, and A. Edelev Collaborative Development and Use of Scientific Applications in Orlando Tools: Integration, Delivery, and Deployment, *Communications in Computer and Information Science*, 2020, vol. 1087, pp. 18–32. doi:10.1007/978-3-030-41005-6_2.
- [12] M. B. Juric, B. Mathew, and P.G. Sarang, Business process execution language for web services: an architect and developer's guide to orchestrating web services using BPEL4WS, Packt Publishing Ltd, 2006.
- [13] F. Biscani and D. Izzo, A parallel global multiobjective framework for optimization: pagmo, *Journal of Open Source Software*, 2020, vol. 5, no. 53, pp. 2338. doi:10.21105/joss.02338.

Разработка системы управления сбором, хранения и контроля Big Data в задачах геоакустики

Алена В. Хрыпченко ¹

¹ *Институт вычислительной математики и математической геофизики СО РАН, проспект Академика Лаврентьева 6, г. Новосибирск, 630090, Россия*

Аннотация

Разработка системы управления базой геоакустических данных направлена на обеспечение эффективного хранения, обработки и доступа к этим данным. С учетом потребностей пользователей в сферах геофизики и геологии, система создается с целью обеспечить оперативный и точный доступ к информации для проведения научных исследований. В данной статье рассмотрим разработку системы для автоматизации управления большими объемами данных в задачах геоакустики с использованием фреймворка Spring, библиотеки визуализации Highcharts и базы данных MySQL. Результаты включают в себя разработанный инструмент для быстрого поиска, анализа и визуализации геоакустических данных, что позволит улучшить качество научных исследований и принимать обоснованные решения. Практическая значимость данной системы проявляется в повышении производительности и качества работы специалистов в области геофизики и геологии, сокращении времени на обработку данных и увеличении эффективности деятельности в области геоакустических исследований.

Ключевые слова

СУБД, геоакустические данные, система, MySQL, Spring, реляционная база данных, нереляционная база данных, Highcharts.

1. Введение

В настоящее время, управление большим объемом данных стало одним из важных аспектов в различных областях науки и технологий. С постоянным увеличением количества информации, ее разнообразием и сложностью, возникает необходимость в эффективной системе управления базами данных (СУБД). Эффективное управление базой данных позволяет хранить, организовывать, обрабатывать и анализировать данные с высокой скоростью и точностью.

В работе будут рассмотрены основные принципы работы СУБД в задачах геоакустики, а также методы и подходы к проектированию и разработке системы управления базой данных. Будет рассмотрена роль языка структурированных запросов SQL в создании и управлении базы данных, а также принципы работы распределенных баз.

Также в работе обсуждается вопрос о выборе наиболее подходящей СУБД для геоакустических задач. Рассматриваются основные характеристики, такие как производительность, масштабируемость, надежность и безопасность, и их роль в выборе подходящей системы.

Технологические аспекты разработки и внедрения базы данных также будут обсуждаться в данной работе. Описываются основные этапы процесса разработки СУБД, включая проектирование базы данных.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia
EMAIL: ahrypchenko@yandex.ru (A. 1)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.24

Практические аспекты внедрения базы данных в реальные системы также являются важной темой статьи. Будут рассмотрены проблемы и решения, связанные с миграцией данных, управлением изменениями, резервным копированием и восстановлением данных, а также безопасностью базы данных.

2. Постановка задачи

Геоакустические данные — это информация, получаемая при помощи геоакустических методов и технологий. Геоакустические данные имеют огромное значение для науки и промышленности по ряду причин.

В научных исследованиях они позволяют узнать больше о структуре земной коры и геологических процессах. Эти данные помогают геологам и геологоразведчикам определить глубину и состав слоев коры, местоположение и характеристики нефтяных и газовых месторождений, а также обнаружить и изучить подземные водные ресурсы и пустоши. Они также помогают исследователям в изучении динамики землетрясений и вулканической активности.

В промышленности геоакустические данные используются для разведки и добычи полезных ископаемых, таких как нефть и газ. Они позволяют определить вероятность наличия месторождения и оптимизировать процессы добычи, уменьшая затраты и увеличивая эффективность. Кроме того, геоакустические данные могут быть использованы при планировании строительства промышленных объектов, таких как нефтяные или газопроводы, позволяя избежать потенциальных проблем, связанных с геологическими условиями.

При работе с большими объемами данных в геоакустике возникают следующие трудности:

1. Ограниченные вычислительные ресурсы: обработка и анализ больших объемов данных требуют значительных вычислительных мощностей. Неправильная организация вычислительного процесса может привести к длительным задержкам и низкой производительности.

2. Затруднения в хранении данных: большие объемы данных требуют высокой емкости хранилища и эффективного способа их организации. Неправильное хранение данных может привести к потере информации или затруднить доступ к необходимым данным.

3. Высокая стоимость обработки данных: обработка больших объемов данных в геоакустике требует специализированного программного обеспечения и оборудования. Это может потребовать значительных финансовых затрат.

4. Сложность анализа данных: большие объемы данных могут содержать различные типы информации, что требует сложного анализа и обработки. Это может вызвать сложности в выделении и интерпретации нужной информации.

5. Проблемы с качеством данных: большие объемы данных могут быть подвержены различным искажениям и шумам. Это может привести к неточным результатам анализа или прогнозированию.

6. Угрозы безопасности данных: большие объемы данных требуют обеспечения их конфиденциальности и защиты от неавторизованного доступа. Недостаточная безопасность данных может привести к утечке информации или нарушению права на конфиденциальность [5].

Для оптимизации геоакустических данных ИВМиМГ было принято решение разработать систему по автоматизации, сбору и хранению Big Data.

3. База данных разрабатываемой системы

Для реализации системы по автоматизации и исследованию геоакустических данных была разработана база данных, состоящей из 23 таблиц. Более детально рассмотреть основные компоненты разработанной системы можно на рисунках ниже.

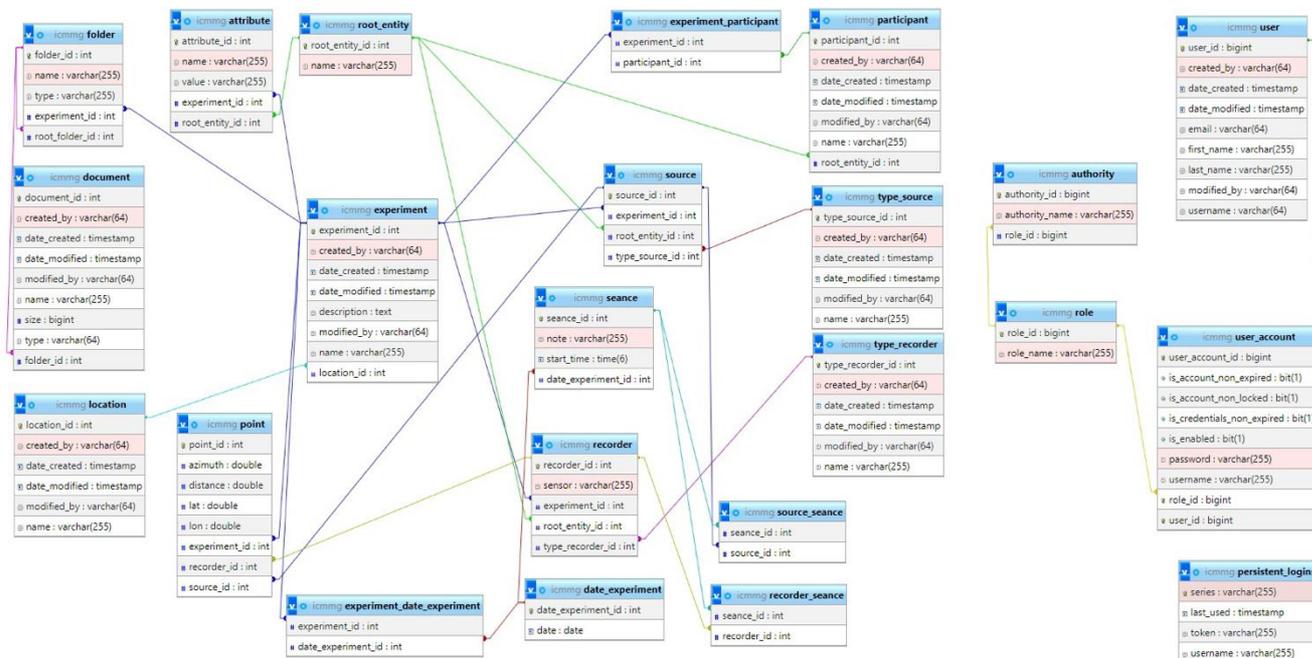


Рисунок 1: База данных для задач геоакустики

Разрабатываемая система построена с использованием Spring Framework на микросервисной архитектуре, представленной на рисунке 2, в которую входят: сервер авторизации. Он представляет из себя единую точку входа для клиентов системы использующий технологию OAuth2.0 для авторизации и Openid Connect для аутентификации. Gateway - для перенаправления запросов клиента к нужному сервису. Цепочка сервисов: реализующие сервисный слой и изменяющие состояние системы. Они связываются с хранилищами данных или выполняющие иную функцию, например, как Email Service рассылки (используется для отправки логов об ошибках на почту администратора). А также Web client для пользователей в браузере.

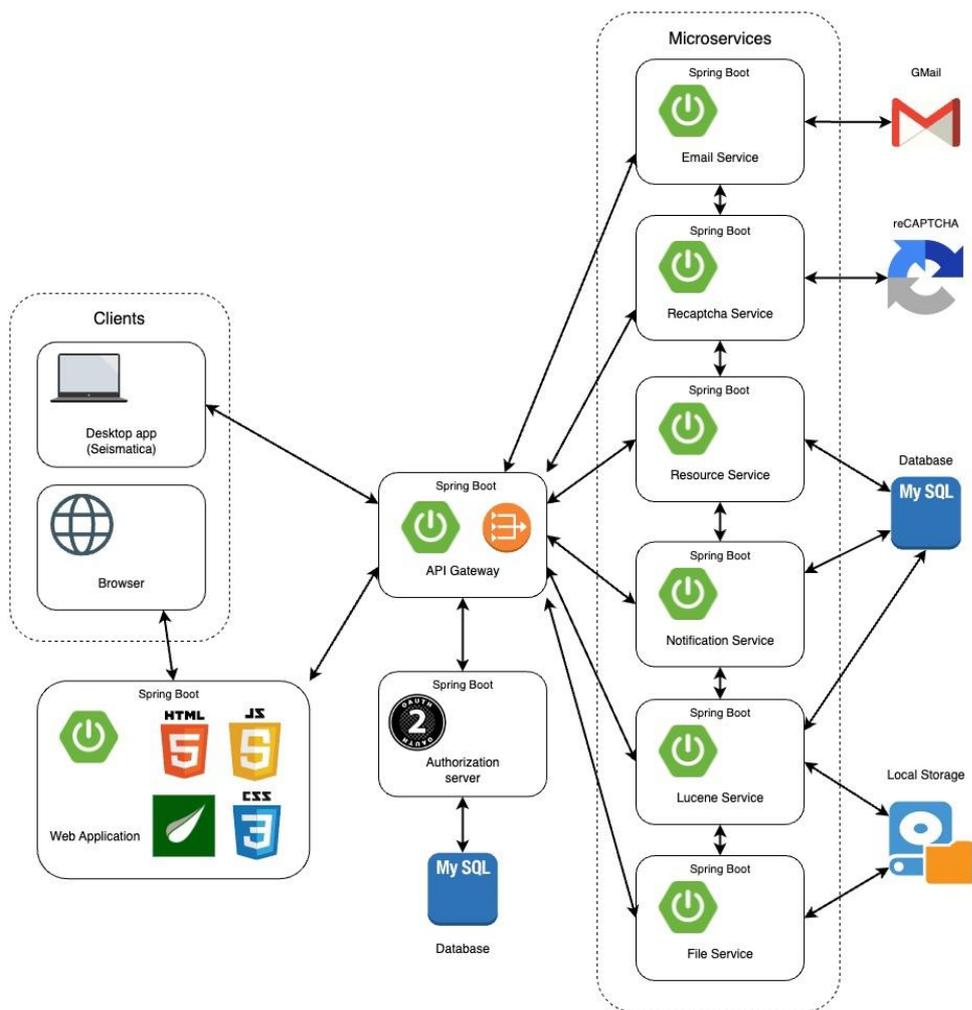


Рисунок 2: Архитектура системы

4. Краткая характеристика результатов

Мной была разработана система по управлению базой данных в задачах геоутики. После добавления системным администратором пользователя в систему, зарегистрированный может войти в систему (рисунок 3). Для безопасности утечки информации была сделана двухфакторная аутентификация.

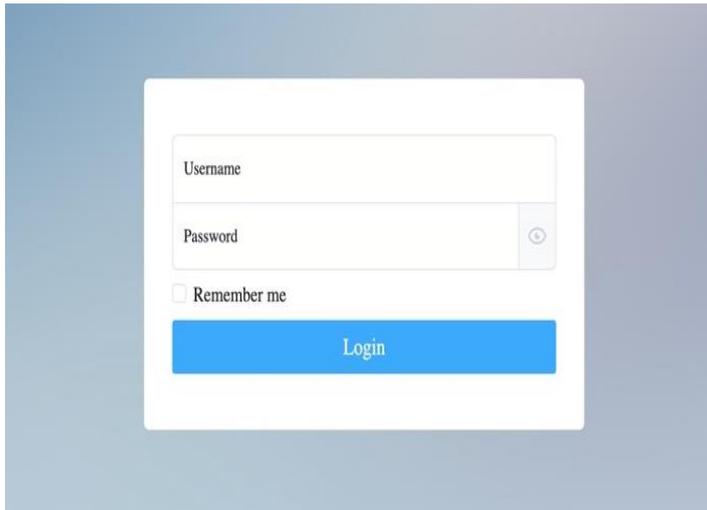


Рисунок 3: Форма входа в систему

Для регистрации нового эксперимента пользователю нужно заполнить форму, где нужно ввести название эксперимента, участников, датчики, местоположение, место проведения, даты, краткое описание эксперимента (рисунок 4).

 A screenshot of a web-based form for registering an experiment. The form is divided into several sections:

- EXPERIMENT:** Contains a dropdown menu with "разрез Ерунак - Крест" selected, another dropdown with "профиль полигон Быстровка - г. Анжеро - Судженск, г. Белово, г. Прокопьевск", and an "Add Location +" button. To the right is a rich text editor with a menu (File, Edit, View, Insert, Format, Tools, Table, Help) and a text area containing Russian text about a seismic experiment from 2002.
- PARTICIPANT:** Features a dropdown menu with "#1 ИВММГ СО РАН" selected and an "Add Participant +" button.
- SOURCE:** Contains three entries. Each entry has a dropdown menu (e.g., "#1 Взрыв"), an "Add Source +" button, and a "Point" section with an "Add Point +" button. The "Point" section includes a table with columns for "#1", "Azimuth", and "Distance".
- RECORDER:** Contains three entries. Each entry has a dropdown menu (e.g., "#1 ПОСА Д"), an "Add Recorder +" button, and a "Point" section with an "Add Point +" button. The "Point" section includes a table with columns for "#1", "Azimuth", and "Distance".

DATE

#1 03.07.2002

Seance

#1 13:32:28

Note

#1 Взрыв #1 РОСА Д

#2 Взрыв #2 РОСА

#3 ВИРС-М

#2 04.07.2002

Seance

#1 15:44:39

Note

#1 Взрыв #1 РОСА Д

#2 Взрыв #2 РОСА

#3 ВИРС-М

Save Cancel File Manager

Рисунок 4: Формы регистрации нового эксперимента

После чего эксперимент можно просмотреть, редактировать и удалить, если на то есть права у пользователя (рисунок 5).

PARTICIPANT

#1 ИВММГ СО РАН

LOCATION

профиль полигон Быстровка – г. Анжеро – Судженск, г. Белово, г. Прокопьевск

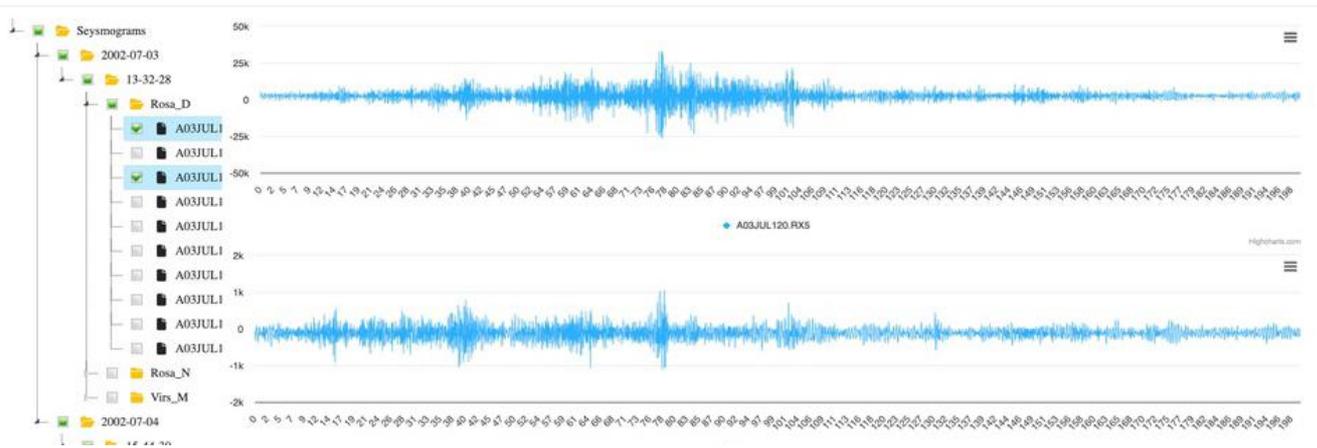
RECORDER

#	Name	Sensor	Point
1	РОСА Д	Lat	Lon
		54.39	83.05
2	РОСА	Lat	Lon
		54.39	83.04
3	ВИРС-М	Lat	Lon
		54.39	83.04

SEANCE

#	Date	Seance
1	03-07-2002	Start time
		13:32:28
		Source №
		Взрыв
		Recorder №
		РОСА Д, РОСА, ВИРС-М
2	04-07-2002	Start time
		15:44:39
		Source №
		Взрыв
		Recorder №
		РОСА Д, РОСА, ВИРС-М

SEISMOGRAM



LES

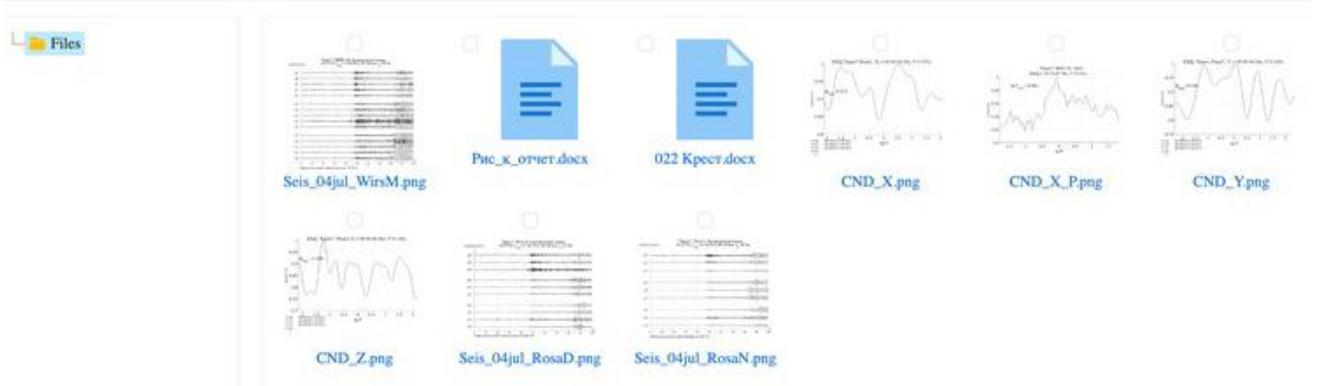


Рисунок 5: Формы просмотра и редактирования эксперимента

5. Заключение

Развитие технологий управления базами данных продолжает оказывать значительное влияние по обработке больших объемов информации. Важно не только следить за новшествами в этой области, но также уделять должное внимание практическим аспектам интеграции полученных решений. Именно проработанный подход к разработке и вдумчивое внедрение СУБД определяют успех её функционирования в долгосрочной перспективе.

Основной целью было предложить комплексный подход к разработке и внедрению систем управления базами данных, который позволял бы эффективно решать задачи хранения, управления и обработки данных. Для достижения этой цели были рассмотрены различные технологии по проектированию базы данных, их преимущества и недостатки. Была разработана система по управлению базой данных в задачах геоактики. Подводя итоги хочется отметить,

что выбор технологии следует осуществлять с учётом конкретных задач и требований заказчиков.

Необходимо анализировать потребности в обработке и хранении данных, а также оценивать доступные ресурсы и ограничения, чтобы выбрать наиболее подходящую технологию.

В целом, статья представляет собой полезное исследование, которое дает читателям широкий обзор различных технологий и практических аспектов при разработке и внедрении систем управления базами данных. Результаты и выводы статьи могут быть полезны для специалистов, занимающихся разработкой и управлением баз данных, а также для научных исследователей в области информационных технологий.

6. Список источников

- [1] Kholodkov K.I., Aleshin I.M., Exact calculation of a posteriori probability distribution with distributed computing systems // *Computer Research and Modeling*, 2015, с. 539–542, 2020.
- [2] Malyshkin V. Active Knowledge, LuNA and Literacy for Oncoming Centuries // *Programming Languages with Applications to Biology and Security: Essays Dedicated to Pierpaolo Degano on the Occasion of His 65th Birthday. Lecture Notes in Computer Science / Bodei C., Ferrari G., Priami C. (ред.). — Springer Cham, 2015. — Т. 9465. — 375 С. — DOI 10.1007/978-3-319-25527-9_19.*
- [3] The sound amplifying forest with emphasis on sounds from wind turbines // Elis Johansson // Department of Civil and Environmental Engineering. Division of Applied Acoustics. Chalmers University of Technology. Sweden. – 2010. – 97 P.
- [4] L.George. Getting Started with Big Data Architecture. <http://blog.cloudera.com/blog/2014/09/getting-started-with-big-data-architecture/>
- [5] D. Laney 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group Inc. File 949. 2001.
- [6] Hu H., Wen Y., Chua T.-S., Li X. Towards scalable systems for big data analytics: a technology tutorial // *IEEE Xplore Digital Library*. 2014. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6842585>

Синтаксический анализатор русского языка на основе данных морфологии: функциональный подход

Евгений Черкашин^{1,3}, Наталия Свердлова² and Елена Марьясова²

¹Институт динамики систем и теории управления СО РАН, Ул. Лермонтова, д. 134, Иркутск, 664033, Россия

²Иркутский научный центр СО РАН, Ул. Лермонтова, д. 134, Иркутск, 664033, Россия

³Институт математики и информационных технологий Иркутского государственного университета, Бульв. Гадарина, д. 20, Иркутск, 6640003, Россия

Аннотация

Задача исследования состоит в создании технологии синтаксического анализа предложений русского языка при помощи обработки морфологических данных слов из корпуса языка АОР, функций-предикатов распознавания соответствия слов по грамматическим (синтаксическим) правилам, управляемого эвристическим переборным алгоритмом, реализованным на компилируемом функциональном языке программирования Haskell.

1. Введение

Задача распознавания синтаксической структуры (РСС) русского и английского текста в документах - одна из часто встречающихся. Требования к качеству РСС предложений варьируется от задачи к задаче. В одних документах требуется объединить строки, формирующие абзац, в предложения, а в других - определить характеристики предложений, а, иногда, и смысл изложенного. Если не требуется полного соответствия грамматике языка, то достаточно применить алгоритм эвристического определения конца предложения, выдающего правильный ответ с некоторой высокой вероятностью 0.95 [1]. Более глубокий вид анализа представляет собой технология link-grammar (грамматика связей) [2], подход достаточно универсальный, и, в целом, независим от структуры языка. Идея link-grammar состоит в представлении модели связей каждого слова в предложении с другим словом при помощи вариантов сцепления. Авторы работы [2] показали, что основным ограничением, наиболее вероятно, имеющим естественные причины возникновения, является тот факт, что получаемый граф связей должен быть планарным. Наборы слов, для которых невозможно построить такой граф не воспринимаются человеком как предложение, несущее смысл. Еще одним направлением РСС является построение дерева синтаксического разбора (ДСР), состоящего из направленных связей)

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ eugeneai@irkn.net (Черкашин); nsverdlova@yandex.ru (Свердлова); mariaselena@yandex.ru (Марьясова)

🆔 0000-0003-2428-2471 (Черкашин); 0000-0002-5315-6266 (Свердлова); 0000-0002-3504-9416 (Марьясова)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (iccs-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.25

на основе грамматических зависимостей (dependency grammar).

Целью данного исследования - реализация подхода к РСС использующего грамматические зависимости, где связи между словами формируются по правилам русского языка. Связь считается возможной, если выполняются ограничения на морфологическое соответствие слов, входящих в связь (правило), и полученный граф связей предложения также является планарным. В отличие от подходов [3] варианты связей в предложении строятся только на основе данных корпуса русского языка, в виде варианта построения нагруженного И-ИЛИ-дерева вариантов распознавания. При внесении дополнительных модификаций данный алгоритм может быть расширен на решение задачи интерпретации дейктических связей между предложениями, например, указание местоимением существительного.

2. Методика представления процесса распознавания синтаксической структуры предложения

Первым этапом проекта является синтаксический анализ предложения. Подход, реализуемый на данном этапе, основан на восходящем распознавании синтаксической структуры по правилам русского языка. В результате строится планарное дерево (набор вариантов таких деревьев) синтаксического разбора, ассоциированное с оценкой вероятности реализации данного разбора. В качестве языка реализации выбран Haskell, позволяющий задавать правила грамматики виде функций-предикатов на основе сопоставления с шаблоном, входной поток так называемых грамем - в виде потенциально бесконечного списка, процесс трансляции - преобразование списка грамем в список синтаксических структур, варианты разбора - списком процессов, и, самое главное, за счет ленивой модели вычисления производить обработки этих потенциально бесконечных структур. Кроме того, Haskell порождает скомпилированный запускаемый модуль. Исходный поток грамем получается из внешнего локального сервиса, реализованного на языке Python и библиотеки Pymorphy2 [4].

Приведем пример задания правил распознавания, представленных в виде функций join, join3 класса Rule.

```
instance Rule GRAM where
  join :: GRAM -> (Gram -> Gram -> Bool,
                  Gram -> Bool, Gram -> Bool)
  join AdjNoun = (adjNounConsist, isAdj, isNoun)
  join NumrNoun = (numrNounConsist, isNumr, isNoun)
  join SubjVerb = (subjVerbConsist, isSubj, isVerb)
  join VerbTranObjAccs = (isAnyRel, isVerbTran, isObjAccs) -- (1)
  join NounNounGent = (isAnyRel, isNoun, isNounGent)
  join Percent = (isAnyRel, isNum100, isPercent)
  join PhoneNumber = (isAnyRel, isWord "+", isPhoneNumber)
  join Sentence = (isAnyRel, hasWall, isSentenceEnd)
  join AdvbVerb = (isAnyRel, isVerb, isAdvb)
```

```

-- . . . . .
join _ = (lfm, lf, lf)
  where lf _ = False
        lfm _ _ = False

join3 :: (GRAM, String) ->
  (Gram -> Gram -> Bool, Gram -> Bool, Gram -> Bool)
join3 (ForJoin, "для") = (isAnyRel, isNoun, isNounGent)
join3 (NounInNoun, "в") = (isAnyRel, isNoun, isNounLoct)
-- . . . . .
join3 _ = (lfm, lf, lf)
  where lf _ = False
        lfm _ _ = False

```

Структура GRAM представляет собой множество грамматических характеристик слов корпуса русского языка АОТ (Автоматическая обработка текста) [5], а также обозначения грамматических связей, все конструкторы являются константами. Функция join ассоциирует конструктор GRAM с тремя предикатами - а) двухаргументный предикат, задающий ограничение на совместное наличие характеристик “склеиваемых” в грамматическую структуру слов, б) два одноаргументных предиката, задающих индивидуальные, независимые характеристики склеиваемых слов. Предикат isAnyRel обозначает, что на слова не задаются совместные ограничения, т.е. для любых входных слов возвращает True.

В тексте примера меткой (1) выделено правило распознавания переходного глагола с существительным в винительном падеже. Здесь не накладывается совместных ограничений. Рассмотрим пример разбора простейшего предложения “мама мыла раму”. Граммемы r morphology2 слов следующие (вывод сокращен):

мама:

1. word= 'мама', 'NOUN,anim,femn sing,nomn', nf='мама', score=1.0,

мыла:

1. word= 'мыла', 'NOUN,inan,neut sing,gent', nf='мыло', score=0.333

2. word= 'мыла', 'VERB,impf,tran femn,sing,past,indc', nf='мыть', score=0.333

3. word= 'мыла', 'NOUN,inan,neut plur,nomn', nf='мыло', score=0.166

4. word= 'мыла', 'NOUN,inan,neut plur,accs', nf='мыло', score=0.166

раму:

1. word= 'раму', 'NOUN,inan,femn sing,accs', nf='рама', score=0.888

2. word= 'раму', 'NOUN,inan,masc,Geox sing,datv', nf='рам', score=0.11

Разбор начинается с первой пары слов предложения. Первые два слова, которые можно объединить - это “мама” и “мыла”. Граммемы слов порождают четыре комбинации, которые фильтруются предикатами правил: производится неэвристический перебор правил, запускаются функции-предикаты. Правила, предикаты которых истинные, задают вариант связи. Оценка связи делается перемножением значения score участвующих в связи слов. В нашем примере только одно правило остается - “подлежащее-глагол” (SubjVerb), где подлежащее должно

быть существительным (NOUN, femn, sing) в именительном падеже (nomn), глагол (VERB) соответствовать роду и числу подлежащего (femn, sing). Формируется связь с оценкой 0.333. Следующее слово - “раму”, оно может быть соединено или с “мыла” или с “мама”. Правило согласования “мыла раму” - переходный (tran) глагол (VERB) с существительным (NOUN) в винительном падеже (accs) (VerbTransObjAccs). В результате получается две последовательные связи в предложении, состоящем из трех слов. Общая оценка разбора - сумма оценок всех сформированных связей, достаточная чтобы можно было сравнить варианты разбора.

Правила, заданные в программе, предполагают прямой порядок слов в предложении (SVO). Если для очередного слова не удается построить связь с предыдущим словом (обратный порядок слов, вводная фраза, отсутствие правила) порождается три альтернативных варианта: а) создание связи общего вида - “слабый” вариант разбора, б) пропуск слова и формирование процесса РСС вводной фразы, в) помещение слова в список слов, расположенных в обратном порядке для последующего продолжения анализа по основному варианту. Оценка связи для варианта (а) равняется 0.1 и уменьшается в 0.7 раза для всех последующих таких связей. В случае (б) строится отдельная фраза.

3. Развитие средств анализа текста

Основная цель проекта – разработать методику машинного перевода научного текста с одного естественного языка на другой, привязанного к заданной предметной области. Одним из требований выступает необходимость реализации системы логического обоснования принятых решений. Процесс перевода включает несколько этапов: 1) анализ синтаксической структуры предложения, 2) выделение в предложении отдельных фраз и связывание их в древовидные синтаксические структуры общего вида (ДСОВ), 3) привязка фраз к концептуальной модели предметной области текста, 4) интерпретация ДСОВ исходного языка в ДСОВ целевого, 5) порождение предложения целевого языка. В качестве исходного языка выбран русский язык, а целевого - английский (и наоборот), как языки представляющие наибольший интерес в части преподавания профессионального перевода.

Работы ученых в области РСС в 80х-90х годах 20 века показали, что качественный перевод, реализуемый без использования машинного обучения, требует, чтобы программа не только анализировала синтаксическую структуру предложения, отображала термины при помощи словаря, но и понимала в некоторой степени суть описанного, потенциально, задавала вопросы с целью уменьшения неопределенности, высказывала несогласие с изложенным. Стандартный подход к моделированию некоторой знаковой системы, включая автоматический перевод, включает задание трех моделей: семантической, синтаксической и модель прагматики. Соответственно, для реализации качественного перевода необходимо формировать модель предметной области до начала реализации перевода, т.е. семантику и прагматику, а также алгоритмы привязки синтаксических структур

к этим моделям.

Модели семантики и прагматики переводчика предполагается задавать при помощи расслоения [6] онтологий, концептуальных моделей, представленных явно соответствующими средствами при помощи формального языка. Каждый слой расслоения представляет некоторый аспект предметной области, между слоями задаются морфизмы, отображающие элементы одной онтологии, более специализированной, на более общую. Слои задаются средствами моделирования онтологий (например, Protege, UML, mind maps), а морфизмы - при помощи правил Prolog и отношений OWL2. Кроме того, слой и совокупность слоев представляется в виде объекта Logtalk, объектно-ориентированной надстройки над языком Prolog. Использование Logtalk позволяет инкапсулировать модели семантики и прагматики в рамках одного объекта, интерпретируемого, в частности, как фрейм Марвина Мински.

4. Варианты использования в обучении профессиональному переводу

Обучение профессиональному переводу научных текстов связано с рядом проблем, частичное решение которых возможно при помощи машинного перевода. Например, в задаче обучения русскоговорящих студентов иностранному языку научного общения, так и иностранных студентов русскому языку, необходимо формирование базиса научной коммуникации. Программы на основе технологий синтаксического анализа, будучи использованными в учебном процессе, способствуют пониманию структуры предложения и его грамматики, в том числе типы предложений, порядок слов, основные синтаксические конструкции и правила их построения, связи между лексическими единицами. В результате повышается эффективность процесса обучения иностранному языку, ускоряется приобретение навыков научного общения, позволяя студентам лучше понимать и овладевать структурой и особенностями изучаемого языка.

Автоматический анализ больших объемов текстов позволяет повышать качество перевода, выявляя языковые закономерности, тенденции и особенности контекста. Данные такого анализа могут использоваться для лингвистических исследований, анализа структуры научных текстов и изучения различий между разными стилями общения. Одно из прикладных направлений - это автоматическое создание рефератов, аннотаций или инструменты для коррекции структуры и грамматики текстов.

На базе разрабатываемых технологий аспиранты и молодые ученые могут создавать собственные языковые продукты, в частности научные статьи, как на русском, так и на иностранном языках, при автоматизированной поддержке обнаружения грамматических ошибок, структурных несоответствий в текстах и не коррелирующих с исходной интенцией конструкций, что, в свою очередь, повышает языковую грамотность.

При работе над научными статьями на иностранном языке синтаксический

анализатор будет эффективным инструментом перевода и адаптации своих исследований. Это позволяет более точно передать на иностранном языке смысл и структуру предложений из русскоязычного оригинального текста, что важно для сохранения точности и качества перевода. При этом обеспечивается грамматическая корректность и логическая последовательность предложений, оценивается и поддерживается стилистическое соответствие текста. Разрабатываемые инструменты могут значительно облегчить процесс создания научной статьи на иностранном языке, повышая профессионализм и качество исследовательских работ молодых ученых и аспирантов.

Исследования в области лингвистики также могут быть поддержаны инструментально, например, в задаче выделения ключевых терминов и специализированной терминологии в научных текстах определенной научной специальности. Здесь важно создать актуальный глоссарий научного дискурса, включающий все релевантные термины и определения для конкретной предметной области. РСС позволяет определить синтаксические связи и отношения между различными терминами в тексте, сформировать полные и точные определения и связи между терминами в глоссарии. В конечном итоге указанные технологии обеспечивают точность, структурирование и организацию информации о терминах и их отношениях, облегчая понимание и обмен информацией в рамках научного институционального дискурса.

Формирование предметной области в сознании носителя научной информации как компонента системы адаптивного обучения является ведущим условием самостоятельного моделирования научно-исследовательской деятельности. Таким образом, полученный набор имманентных корпусу научных текстов свойств будет использован для создания нового языкового продукта, который представляет собой языковой инвариант - готовую семантико-синтаксическую основу для описания процесса и результатов исследования в лингвистике.

5. Заключение

В статье представлены результаты начального этапа разработки предположительно однопроходного синтаксического анализатора предложений и фраз русского языка, адаптирующего базовые принципы алгоритма link-grammar [2] и грамматик зависимостей к вычислительной модели функционального языка программирования Haskell. Входной поток слов предложения (фразы), охарактеризованных так называемыми граммемами, представляется как бесконечный список структур. Данные структуры преобразуются в список грамматических связей, одновременно формирующих дерево грамматического разбора. Варианты построения синтаксического разбора оцениваются. Оценка позволяет задавать предпочтения того или иного варианта, устанавливать приоритет процесса дальнейшего разбора.

На данном этапе проекта производится уточнение структуры алгоритма, анализ литературы по грамматике русского языка, формализация правил языка, реализация библиотеки-аналога r morphology2 [4] средствами Haskell с ее адаптацией

к режиму ленивых вычислений. Реализация технологии направлена на решение задачи распознавания текста в неразмеченных документах формата PDF или результатов сканирования/распознавания изображений.

Следующие этапы исследования связаны с построением моделей семантики и прагматики в виде полисистемного расслоения [6], частично представленного в проекте OpenCyc [7], а также правил распознавания отношений в дереве синтаксического разбора, реализованных, например в [8].

6. Благодарности

Исследование проведено при поддержке Министерства науки и высшего образования Российской Федерации, проект ИИЦ СО РАН «Лингвосомиотическая гетерогенность научной картины мира: теоретическое и лингводидактическое описание» (FWSN-2022-0001), № государственной регистрации, и ИДСТУ СО РАН «Методы и технологии облачной сервис-ориентированной цифровой платформы сбора, хранения и обработки больших объёмов разноформатных междисциплинарных данных и знаний, основанные на применении искусственного интеллекта, модельно-управляемого подхода и машинного обучения», (FWEW-2021-0005).

Список литературы

- [1] Sentence boundary disambiguation. “Doing Things with Words, Part Two: Sentence Boundary Detection”, URL:https://en.wikipedia.org/wiki/Sentence_boundary_disambiguation#cite_note-2 (access date: 26.05.2024)
- [2] D. Grinberg, J. Lafferty, D. Sleator. A robust parsing algorithm for link grammars. Carnegie Mellon University Computer Science technical report CMU-CS-95-125, and Proceedings of the Fourth International Workshop on Parsing Technologies, Prague, September, 1995. URL:<https://www.cs.cmu.edu/afs/cs.cmu.edu/project/link/pub/www/papers/ps/tr95-125.pdf> (access date: 26.05.2024)
- [3] Ru-eval: оценка методов автоматического анализа текстов. URL:<https://ru-eval.github.io/syntax.html> (access date: 26.05.2024)
- [4] Korobov M.: Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts, 2015, pp 320-332.
- [5] Автоматическая Обработка Текста. URL:<http://aot.ru/> (access date: 26.05.2024)
- [6] А. Черкашин, Полисистемный анализ и синтез : Прил. в географии / А. К. Черкашин; Отв. ред. В. С. Михеев;. - Новосибирск : Наука : Сиб. предприятие, 1997. - 499 с
- [7] The OpenCyc Platform. URL:<https://github.com/asanchez75/opencyc> (access date: 26.05.2024)
- [8] J Apresian, I Boguslavsky, L Iomdin, et al. ETAP-3 linguistic processor: A full-fledged NLP implementation of the MTT. First International Conference on Meaning-Text Theory (MTT'2003), pp. 279-288

Semantic Column Annotation for Russian Tables Using Multilingual BERT Language Model

Kirill V. Tobola^{1,2}, Nikita O. Dorodnykh²

¹*Institute of Mathematics and Information Technologies, Irkutsk State University*

²*Matrosov Institute for System Dynamics and Control Theory, Siberian Branch of the Russian Academy of Sciences (ISDCT SB RAS)*

Abstract

Tables are widely used for organizing, presenting, and storing data. Automatically extracting information from these tables is challenging as they often have arbitrary and non-uniform layouts. Recent research has demonstrated significant progress in addressing various challenges in automatic table understanding. But, these are often limited to languages such as English. In this paper, we propose a new RuTaBERT framework, based on a pre-trained language model, for semantic annotation of columns in Russian-language tables. We utilized a preprocessed large-scale corpus of Russian-language web tables, extracted from Wikipedia pages, as training data to fine-tune the language model. Experimental results show that RuTaBERT establishes new state-of-the-art performance on new benchmark with additional language for the column type prediction task with micro F1 score of up to 95%.

Keywords

Semantic table interpretation, Column type annotation, BERT, Pre-trained language model, Knowledge graph, Tables, Russian language

1. Introduction

This paper focuses on improving the comprehension of semantic table structures, particularly for tables written in Russian. We suggest a framework called RuTaBERT, which employs pre-trained language models to enhance semantic column annotation. RuTaBERT is built upon the multilingual BERT model [1] and leverages local table context to boost annotation accuracy. Additionally, we develop a method for automatically labeling source tabular datasets based on table headers. The RuTaBERT codebase and trained models are publicly available, allowing for their application in various systems for table column annotating and performance evaluation.

Overall, the paper offers a significant contribution to the field of table understanding and semantic annotation, with a focus on Russian-language tables. The proposed framework and methods have the potential to increase the efficiency and accuracy of table analysis and knowledge discovery in various domains.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ kirilltobola@gmail.com (K. V. Tobola); nikidorny@icc.ru (N. O. Dorodnykh)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (icc-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.26

2. Methodology

2.1. Problem statement

A table is a 2D array of rows and columns, where each cell can hold various types of data. In this approach, vertical tables are used as input data, where each column may have a header. The goal is to predict the semantic type of each column, such as "Book" or "Publication Date", rather than standard data types like string or integer. To achieve this, a set of 356 semantic types based on the DBpedia knowledge graph has been created, including classes, data properties, and object properties, and translated into Russian. The task is to annotate Russian-language tables with these semantic types.

Figure 1 illustrates an example of solving the CTA task for the input table.

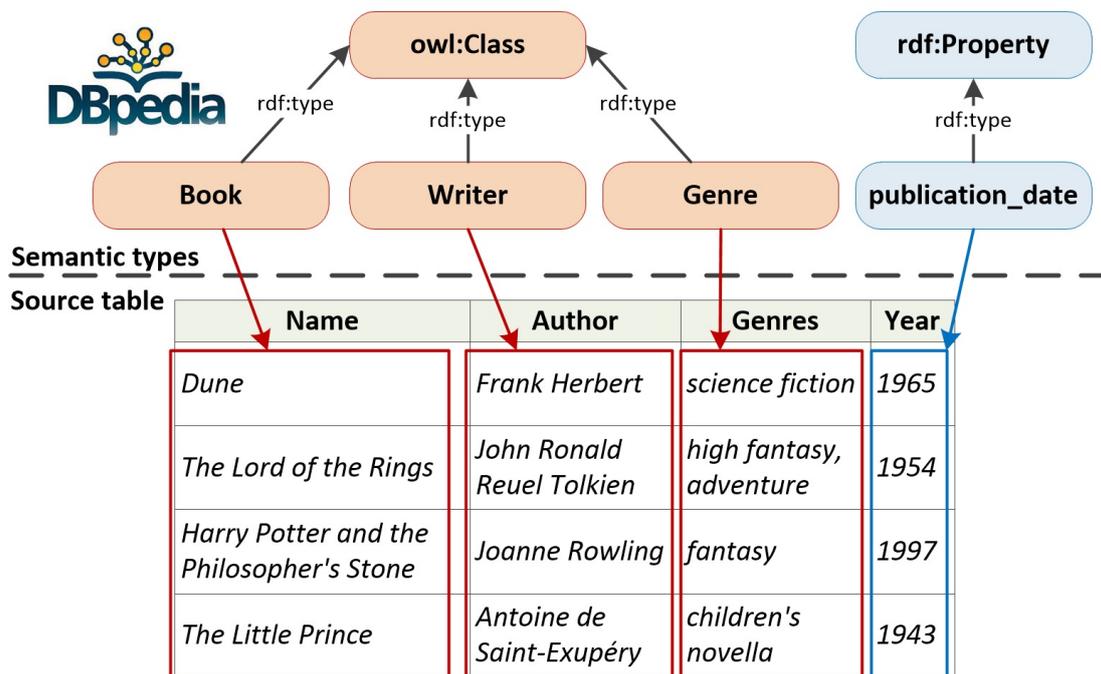


Figure 1: Example of column type annotation problem.

2.2. Model architecture

We suggest a framework called RuTaBERT that fine-tunes the multilingual BERT model for sequence classification tasks, such as column labeling in tables. We also propose two context-aware table serialization methods, multi-column serialization and neighboring column serialization, to convert tables into text sequences for input to the BERT model.

1) *Multi-column serialization* based on the approach [2] and represents data records as follows:

$$S_{mc}(T) = [CLS]tks^1 \dots [CLS]tks^n [SEP]$$

where [CLS] marks the start of a sequence; [SEP] marks the end; tk_i is the tokens for i-column.

In this case, all cell values of a i-column are tokenized: $ci = v_1, \dots, v_m$. This method inserts a [CLS] token for each column, allowing RuTaBERT framework to predict a label for each [CLS] token.

2) *Neighboring column serialization* uses neighboring columns as local context:

$$S_{nc}(T) = [CLS]tk_s^{trg}[SEP]tk_s^{nb}[SEP]$$

where tk_s^{trg} represents the tokens of the target column; tk_s^{nb} represents the tokens for all non-target neighboring columns. Here, only one [CLS] token for the target column is used for prediction, and [SEP] tokens separate non-target columns.

3. Results and discussion

We trained RuTaBERT for 10 epochs with a batch size of 32 using both serialization methods on Russian Web Tables (RWT) [3] dataset. The preprocessed tables from the RWT corpus were split into training, validation, and test datasets with standard proportions: 70%, 20%, and 10%, respectively.

We evaluate RuTaBERT performance on the column type annotation (CTA) task using two serialization methods, multi-column and neighboring column, and compare it to a baseline method based on lexical matching between table header names and semantic types. Table 1 represents the experimental evaluation of RuTaBERT on the CTA task using the RWT dataset.

Table 1

The experimental evaluation for RuTaBERT

Method	Micro F1	Macro F1	Weighted F1
baseline	0.881	0.693	0.902
multi-column	0.952	0.861	0.951
neighboring	0.953	0.856	0.953

According to the results, both serialization techniques have similar performance and achieve better results than the baseline, indicating the framework’s potential for semantic annotation of Russian-language tables. However, the models face challenges related to data sparsity in the table corpus, which is evident in the imbalanced distribution of semantic types. This affects the models’ ability to capture sufficient signals for minority class semantic types.

Acknowledgments

The reported study was supported by the Ministry of Science and Higher Education of the Russian Federation (№ 1023110300006-9).

References

- [1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. [arXiv:1810.04805](https://arxiv.org/abs/1810.04805).
- [2] Y. Suhara, J. Li, Y. Li, D. Zhang, c. Demiralp, C. Chen, W.-C. Tan, Annotating Columns with Pre-trained Language Models, Association for Computing Machinery, 2022. URL: <https://doi.org/10.1145/3514221.3517906>.
- [3] P. Fedorov, A. Mironov, G. Chernishev, Russian web tables: A public corpus of web tables for russian language based on wikipedia, 2022. [arXiv:2210.06353](https://arxiv.org/abs/2210.06353).

Сегментация изображений документов при помощи глубоких нейронных сетей

Андрющенко В.С.¹, Михайлов А.А.^{1,2}

¹Институт динамики систем и теории управления им. В.М. Матросова Сибирского отделения Российской академии наук, 664033, Иркутск, ул. Лермонтова, д. 134, Российская Федерация

²Институт математики и информационных технологий ИГУ, 664003, Иркутск, бульвар Гагарина, 20, Российская Федерация

Аннотация

В статье рассматривается задача автоматического извлечения заголовков из электронных документов. Для решения этой задачи предлагается подход на основе моделей сегментации документов. В ходе исследования была проверена гипотеза о том, что преобразование изображений поможет улучшить качество модели, были собраны и подготовлены данные, а также произведено дообучение моделей нейронной сети.

Keywords

Обработка неструктурированных данных, Анализ электронных документов, Обнаружение заголовков, Глубокое обучение, Методы обработки изображений

1. Введение

Анализ электронных документов является важной задачей в области обработки неструктурированных данных. Он включает в себя распознавание текста, выявление ключевых данных, классификацию, оценку достоверности и соответствия нормативным требованиям. Анализ может использоваться для автоматизации документооборота, повышения эффективности обработки информации и поддержки принятия решений на основе полученных данных. [1]

Одной из задач анализа электронных документов является обнаружение заголовков. Обнаружение заголовков позволяет разбить текст на отдельные физические блоки, что упрощает дальнейший анализ документа. [2] Цель работы – автоматическое извлечение заголовков из электронных документов.

2. Методы и материалы

Для достижения поставленной цели использовались методы на основе глубокого обучения. Используя библиотеку TensorFlow [3], производилось дообучение модели ней-

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ vitya.andr.02@mail.ru (В.С.); mikhailov@icc.ru (А.А.)

ORCID 0009-0001-3404-2044 (В.С.)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (iccs-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.27

ронной сети ResNet 50. [4] Для того, чтобы собрать новый набор данных, использовался инструмент Tagged PDF¹.

3. Результаты

На данном этапе работы гипотеза о том, что преобразования изображений документов поможет улучшить точность модели не подтвердилась, преобразования не дали положительного результата. Был создан новый набор данных, с использованием инструмента: Tagged PDF и обучена модель, которая показала средние результаты при тестировании.

4. Выводы

Получены промежуточные результаты: создан новый набор данных и обучена модель, однако результаты модели при тестировании были лишь средними. Это подчёркивает важность продолжения исследований и оптимизации подходов. Неподтверждение гипотезы о положительном влиянии преобразования изображений на точность модели свидетельствует о необходимости дальнейшего изучения и экспериментирования с различными методами предварительной обработки данных.

Дальнейшие исследования должны сосредоточиться на улучшении качества данных, расширении набора данных, оптимизации архитектуры моделей и применении более сложных методов обработки изображений.

5. Заключение

Полученные результаты подчёркивают сложность задачи автоматического извлечения заголовков из электронных документов и необходимость использования комплексного подхода, который включает в себя не только глубокое обучение, но и тщательное предобучение и подготовку данных.

Список литературы

- [1] Томашевская, Яковлев, Способы обработки неструктурированных данных, Russian Technological Journal 9 (2021) 7–17.
- [2] H. Kim, S. Park, S. Yang, Y. Jung, Cilab@ fintoc-2021 shared task: Title detection and table of content extraction for financial document, in: Proceedings of the 3rd Financial Narrative Processing Workshop, 2021, pp. 106–110.
- [3] D. TensorFlow, Tensorflow, Site oficial (2018).
- [4] B. Koonce, B. Koonce, Resnet 50, Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization (2021) 63–72.

¹https://github.com/sunveil/tagged_pdf/tree/master

Методы и средства реализации моделей алгоритмов анализа актинометрических показателей

Дмитрий Карамов¹, Евгений Юмашев¹

¹ Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

Аннотация

В статье представлена реализация WPS-сервиса для моделирования актинометрических показателей для любых локаций. Сервис основывается на модифицированной математической модели Iqbal-Kasten/Czeplak, которая позволяет получать детализированные массивы данных по солнечной радиации.

При реализации WPS-сервиса использовались специализированные библиотеки numpy, scipy, ruwps, os, ruwps-flask, которые позволяют оптимизировать алгоритмы рассматриваемой модели.

Результаты, полученные в ходе данного исследования, позволили значительно уменьшить время расчета, формирования массивов данных и их визуализации. Кроме того, характер полученных функций солнечной радиации говорит об адекватности рассматриваемой модели и WPS-сервиса реализующего её.

Ключевые слова

Солнечная радиация, актинометрические показатели, WPS-сервис, Python, FM 12 Супор, METAR

1. Введение

Актинометрические показатели такие как прямая, рассеянная и суммарная солнечная радиации являются важными показателями для различных областей деятельности человека. Они могут использоваться в строительстве, энергетике, сельском хозяйстве и многих других сферах. Наличие достоверных актинометрических данных является важным фактором, влияющим на конечные результаты прикладных и научно-исследовательских работ [1]. В частности, при решении различных задач в области энергетики мощность солнечной электростанции имеет прямую зависимость от уровня инсоляции. Следовательно, в случае решения задачи выбора оптимальной мощности фотоэлектрической системы необходимо располагать подробным и точным массивом данных [2].

Цель данной статьи состоит в демонстрации подхода для моделирования солнечной радиации и реализующей её открытого WPS-сервиса.

2. Постановка задачи разработки онлайн WPS-сервиса для расчета солнечной радиации

Постановка задачи разработки онлайн сервиса для расчета суммарной солнечной радиации основывается на следующих положениях:

1. Открытый WPS-сервиса;

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL dmitriy.karamov@mail.ru (A. 1); yumashevgeny@mail.ru (A. 2)

ORCID: 0000-0001-5360-4826 (A. 1); 0009-0003-8736-0155 (A. 2)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.28

2. Возможность задавать любую локацию в зависимости от географических координат местности;
3. Автоматически формировать исходную климатическую информацию необходимую для моделирования актинометрических показателей;
4. Использовать данные необходимые для описания динамики масс в средних и нижних слоях атмосферы;
5. Формировать данные актинометрических показателей для типичного, оптимистичного и пессимистичного метеорологического года.

Реализация основных положений данной постановки позволит получить сервис для формирования актинометрических данных для любой локации на основании ретроспективных природно-климатических данных формата FM 12 Synop и METAR. Также используются данные с открытых источников об спектральном составе области видимого света, наличия капель воды и водяного пара в атмосфере, а также озонового слоя.

Учитывая имеющуюся постановку, а также ряд факторов, описывающих требования к математической модели реализуемой в рамках сервиса было решено использовать модифицированную модель Iqbal-Kasten/Czeplak [3].

3. Краткое описание математической модели

За основу была взята модифицированная математическая модель Iqbal-Kasten/Czeplak которая хорошо себя зарекомендовала при решении различных задач в области использования фотоэлектрических систем. В частности, при решении задач развития и функционирования солнечных электростанций.

Укрупненно модель включает в себя три основных блока:

1. Блок подготовки исходных климатических массивов необходимых для моделирования актинометрических показателей.
2. Блок моделирования солнечной радиации для рассматриваемой локации относительно каждого часа расчетного периода по ретроспективным многолетним массивам природно-климатической информации. При этом учитываются действительный часовой пояс территории формата UTC.
3. Блок формирования результатов моделирования прямой, рассеянной и суммарной солнечной радиации с определением типичного, пессимистичного и оптимистичного метеорологического года.

Исходной информацией для математической модели солнечной радиации Iqbal-Kasten/Czeplak являются массивы данных формата FM 12 Synop и METAR [4].

FM 12 SYNOP – это код для оперативной передачи данных приземных гидрометеорологических наблюдений с сети станций гидрометслужбы, расположенных на суше (включая береговые станции). METAR – авиационный метеорологический код для передачи сводок о фактической погоде на аэродроме. Стоит отметить, что в кодах FM 12 SYNOP и METAR соблюдается строгий порядок следования информации. Многолетние метеорологические ряды являются детализированной информацией. При этом правильная обработка данных рядов позволит максимально точно описывать поведение природно-климатических показателей на рассматриваемой территории.

Международные коды FM 12 Synop и METAR имеют одинаковые показатели приборных измерений, а именно: скорость и направление ветра, температура воздуха, давление, влажность и отдельно качественная оценка общей облачности. Под функциональной зависимостью вида $f(t)$ подразумевается значение показателя f в момент времени t . Шаг изменения t в данной случае составляет один час.

Вектор приборных измерений $A_m^r(t)$ в момент времени t года τ расчетного периода, включающий качественную оценку общей облачности в процентах, имеет следующий вид:

$$A_m^r(t) = [v_{wind}(t), T_{air}(t), p_{air}(t), \phi_{\%}(t), c(t)]. \quad (1)$$

Стоит отметить, что код FM 12 Synop содержит более детальную информацию качественной оценки облачности, состава и высоты образования облаков, процент облаков

плохой погоды, облака нижнего яруса (слоисто-кучевые, слоистые, слоисто-дождевые, кучево-дождевые), облака среднего яруса (высококучевые, кучевые, высокослоистые), облака верхнего яруса (перистые, перисто-кучевые и перисто-слоистые облака). В международном коде METAR качественная оценка облачности послойно записана в вектор-столбце, что значительно усложняет их обработку за счет построчного сравнения эталонных (словесных) значений массива относительно каждого временного интервала τ -го года расчетного периода.

Вектор качественной оценки облачности $B_m^r(t)$ в момент времени t года τ расчетного периода для кода FM 12 Synop

$$B^r(t) = [C_l(t), C_m(t), C_h(t), b_{\%}(t)]. \quad (2)$$

Матрица приборных измерений $A_m^r(t)$ и матрица качественной оценки облачности $B^r(t)$ объединяются в единую матрицу природно-климатических показателей, где каждая строка есть вектор, характеризующий фактические значения параметров окружающей среды в определенном момент времени. Таким образом, полученная объединенная матрица описывает природно-климатические показатели с дискретным шагом один час.

Также используется информация об изменении вертикальной толщины озонового слоя, наличии капель воды и пара в облаках. Данные показатели находятся в свободном доступе в электронных архивах World Ozone and Ultraviolet Radiation Data Centre (WOURDC) и Aerosol RObotic NETwork (AERONET). Отметим, что данные показатели необходимы для более точного определения актинометрических характеристик рассматриваемой территории.

С подробным математическим описанием обработки многолетних метеорологических рядов с кодами FM 12 SYNOP и METAR их интеграции в системные энергетические исследования можно ознакомиться в [5]. На рисунке 1 показана блок-схема алгоритма работы математической модели актинометрических показателей.

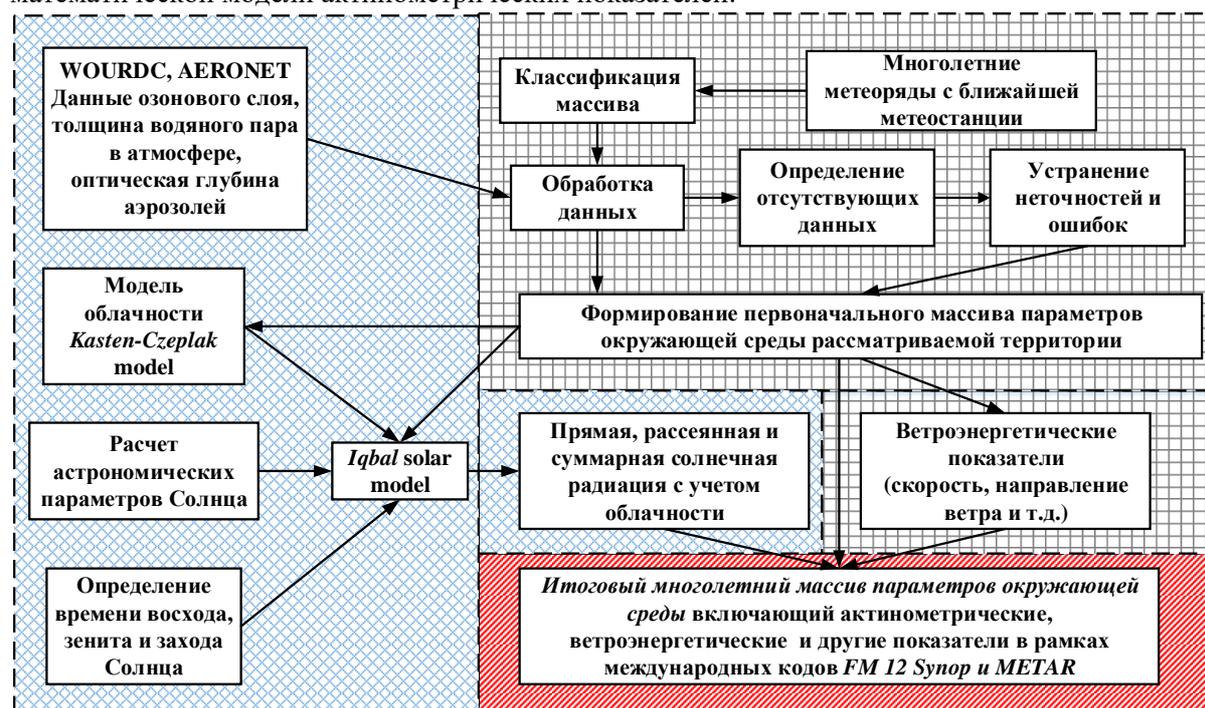


Рисунок 1: Блок-схема математической модели Iqbal-Kasten/Czeplak

4. Разработка WPS-сервиса расчета солнечной радиации

Алгоритмы для расчета показателей солнечной радиации написаны в программной среде MATLAB. Основные трудности использования этих алгоритмов заключаются в неудобстве вывода полученных показателей, а также низкой скорости их выполнения (примерно 40 минут) в зависимости от размерности исходного массива климатической информации и количества лет расчетного периода.

Для решения данных проблем были реализованы сервис-ориентированные методы в виде WPS-сервиса (Web Processing Service). Схема работы WPS-сервиса представлена на рисунке 2.

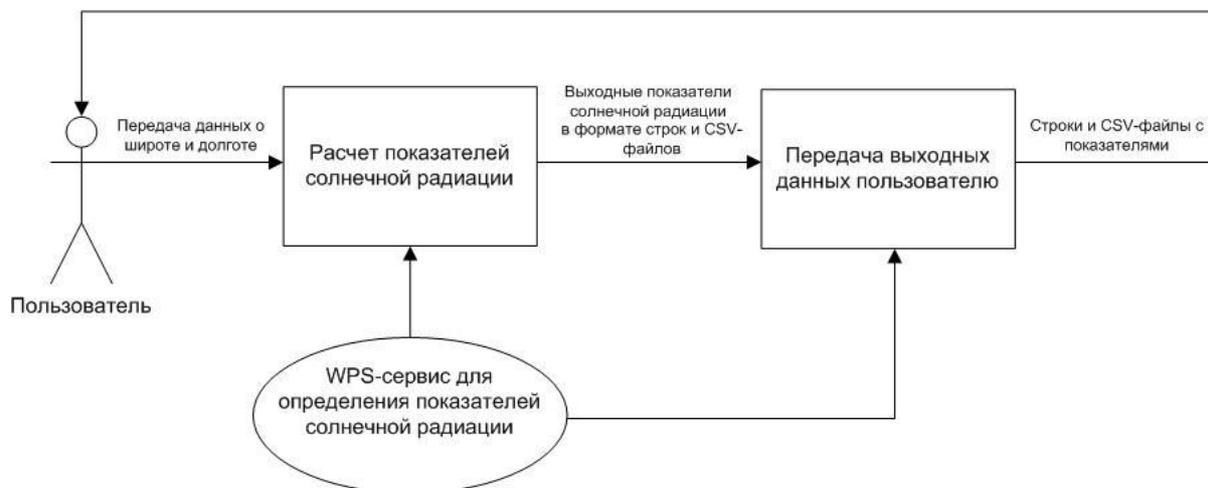


Рисунок 2: Схема работы WPS-сервиса

Пользователь подает на вход в WPS-сервис параметры географической широты и долготы местности, по которым он планирует получить массивы актинометрических показателей. На основе подготовленных данных о фактической облачности, составе облаков, температуре воздуха, давлению, влажности и других параметров, которые были получены с ближайшей метеостанции, а также данных о номере расчетного года, проводится моделирование показателей суммарной солнечной радиации и ее составляющих, а именно: угол возвышения солнца над горизонтом, среднемесячные суммарная, прямая и рассеянная солнечные радиации, их среднечасовые показатели, а также среднегодовые значения.

Для этого были реализованы методы оценки облачности нижнего, среднего и верхнего уровней, метод расчета общего процента облачности и другие параметры. В результате получают показатели суммы, представленные в формате строк, а временные ряды остальных показателей – в формате CSV.

WPS-сервис был разработан при помощи языка программирования Python. Для реализации расчетной части были применены библиотеки `numpy` и `scipy`. Для создания непосредственно самого WPS-сервиса были использованы такие библиотеки, как `ruwps` и `os`, а тестирование происходило при помощи `ruwps-flask`.

В результате написания WPS-сервиса были решены ряд задач с выводом полученных показателей солнечной радиации в удобный для представления формате, а также сокращение времени выполнения алгоритмов до 2 минут. На рисунке 3 показаны результаты моделирования солнечной радиации за 10 лет расчетного периода для территории западной Сибири с 2013 по 2023 годы.

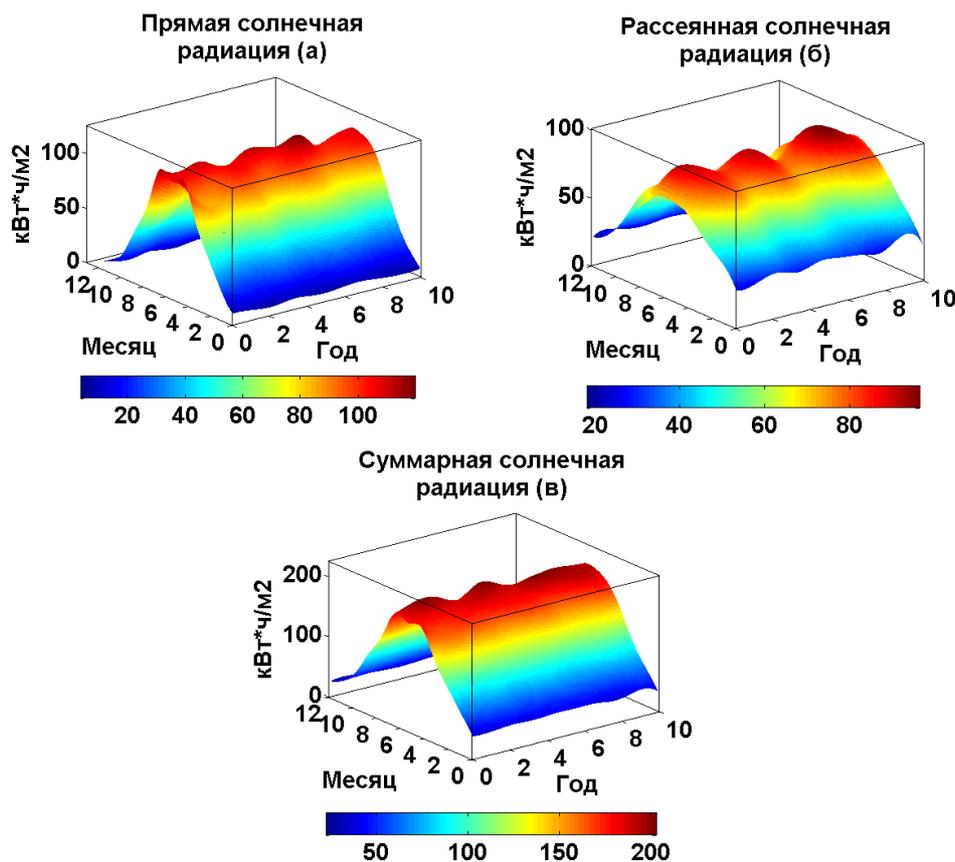


Рисунок 3: Результаты моделирования солнечной радиации для территории западной Сибири

Результаты, полученные при использовании WPS-сервиса, имеют характерный вид для рассматриваемого региона с максимальными значениями в весенне-летний периоды.

5. Заключение

В рамках данной статьи были представлены изначальная постановка исследования, математическая модель Iqbal-Kasten/Czeplak и алгоритмы, реализующие их в виде WPS-сервиса открытого доступа для моделирования солнечной радиации по ретроспективным данным для рассматриваемых локаций.

Использование библиотек numpy, scipy, ruwps, os, ruwps-flask позволило наиболее эффективно реализовать основные алгоритмы, описывающие работу математической модели солнечной радиации.

Дальнейшие исследования будут посвящены развитию реализованных моделей и алгоритмов, а также решению следующих задач:

- уменьшение времени работы алгоритмов для вычисления показателей солнечной радиации путем оптимизации методов расчета и алгоритмов.
- Формирование блока выявления аномальных значений, а также заполнение отсутствующих данных за счет использования методов искусственного интеллекта и других средств в этой области.
- внедрение разработанного сервиса в геопортал ИДСТУ СО РАН.

6. Благодарности

Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды», , рег. № 124052100088-3).

7. Список источников

- [1] Лукутин, Б. В. Оптимизация энергетических балансов фотоэлектрической станции с электрохимическим и тепловым аккумулярованием солнечной энергии / Б. В. Лукутин, А. М. Каррар Хамид // Известия высших учебных заведений. Проблемы энергетики. – 2022. – Т. 24, № 2. – С. 3-13.
- [2] Лукутин, Б. В. Критерии вовлечения ВИЭ в повышение энергетической безопасности изолированных труднодоступных территорий Севера и Арктической зоны РФ / Б. В. Лукутин, В. Р. Киушкина, И. С. Иванов // Журнал Сибирского федерального университета. Серия: Техника и технологии. – 2021. – Т. 14, № 5. – С. 507-519.
- [3] L.T. Wong, W.K. Chow. Solar radiation model / L.T. Wong // Applied Energy. – 2001. – V. 69. – P. 191-224.
- [4] Карамов, Д. Н. Формирование исходных метеорологических массивов с использованием многолетних рядов FM 12 Synop и METAR в системных энергетических исследованиях / Д. Н. Карамов // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2018. – Т. 329, № 1. – С. 69-88.
- [5] Карамов, Д. Н. Математическое моделирование солнечной радиации с использованием многолетних метеорологических рядов, находящихся в открытом доступе / Д. Н. Карамов // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2017. – Т. 328, № 6. – С. 28-37.

Классификация блоков изображения на основе графового представления выделенных слов с использованием нейронных сетей

Даниил Копылов^{1,2}, Андрей Михайлов^{1,2}

¹Институт динамики систем и теории управления им. В.М. Матросова Сибирского отделения Российской академии наук, 664033, Иркутск, ул. Лермонтова, д. 134, Российская Федерация

²Институт системного программирования им. В.П. Иванникова Российской академии наук, 109004, г. Москва, ул. Александра Солженицына, д. 25, Российская Федерация

Аннотация

В данной работе предложен метод классификации структурных элементов текста на основе графового представления этих элементов. Метод использует граф, построенный по словам сегмента, и его статистические характеристики. Точность метода составляет 0.793.

Ключевые слова

Классификация сегментов изображения, анализ макета страницы, представление сегментов

1. Введение

В современном мире информация хранится преимущественно в цифровом виде, значительная часть которой представлена сканированными документами. Обработка такой информации требует распознавания текста и восстановления его логической структуры.

Задача детекции структурных блоков текста (заголовки, основной текст, таблицы) не новая. Одной из ранних работ в этой области является [1]. На сегодняшний день наиболее перспективные решения в этой сфере представлены в работах [2] и [3].

В этих работах для решения задачи используются нейронные сети, основанные на архитектуре Transformer. Однако такие модели, как правило, являются достаточно ресурсоемкими и не способны быстро обрабатывать большие объемы документов, что становится проблемой при работе с тысячами или миллионами файлов.

В качестве альтернативы Transformer-архитектурам в задачах обработки документов набирают популярность подходы, основанные на графовых нейронных сетях. Одним из примеров является работа [4]. Графовые сети уже применялись для решения отдельных задач в этой области, например, сегментации [5].

В данном исследовании мы предлагаем использовать графовый подход для классификации структурных блоков текста. Помимо того, что мы не используем стандартные

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ it-daniil@yandex.ru (Копылов); mikhailov@icc.ru (Михайлов)

🆔 0009-0000-6348-4004 (Копылов); 0000-0003-4057-4511 (Михайлов)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (icc-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.29

графовые сети, мы также строим граф по-другому. В отличие от большинства существующих решений, где графы строятся из участков изображения [6] или строк текста [5], мы предлагаем использовать графы, построенные на основе слов.

Такой подход имеет ряд преимуществ. Во-первых, использование слов естественным образом вписывается в процесс обработки документов, так как на выходе всегда требуется получить текст, а значит его в любом случае необходимо детектировать. Во-вторых, не требуется получать дополнительные элементы из документа, что экономит время и ресурсы при обработке. В-третьих, слова, как правило, распознаются более четко, чем строки, и даже если они сливаются в одно слово, это происходит в рамках одного и того же структурного блока.

Предложенный подход позволит повысить эффективность и скорость обработки больших объемов текстовых документов, что имеет большое значение для различных областей, таких как цифровизация архивов, автоматизация обработки юридической документации и т.д.

2. Предлагаемый подход

В качестве набора данных использовались сегменты текста из PubLayNet [7]. Оттуда было отобрано 10% изображений, которые затем были сбалансированы и разделены на три набора: обучающий (4386 сегментов), проверочный (1096 сегментов) и тестовый (1370 сегментов).

Для распознавания слов в изображениях использовался OCR Tesseract [8]. Было обнаружено, что для повышения качества распознавания необходимо увеличить разрешение изображений (например, разрешение 100 первых изображений из PubLayNet было увеличено в 4 раза, что привело к увеличению среднего количества распознанных слов с 539 до 681).

Граф строится так, чтобы каждое слово имело ссылки на четырех соседей: 2 по строке (слева/справа) и 2 по абзацу (сверху/снизу). В случае границ, слово может ссылаться само на себя (петля). Для ускорения построения: изображение разбивается на секторы, число которых пропорционально числу слов. Поиск соседа начинается в секторе, где находится слово, а затем в соседних секторах (поиск в глубину). Ограничение по глубине поиска позволяет сократить время поиска и избавиться заведомо лишних связей.

После граф обогащается признаками. Условно их можно поделить на признаки ребер и узлов. Для ребер рассчитывается длина и косинус угла наклона ребра к горизонту. Для узлов высота блока слова и оценка "жирности" (отношение площади к периметру области).

Граф напрямую не может быть передан в MLP-модель. Поэтому он преобразуется в вектор, который представляет собой конкатенацию векторов для каждого признака. Вектора признаков - это интервальные ряды распределения значений признаков. Длина векторов подбиралась экспериментально.

Модель представляет собой MLP с 3 слоями:

1. Первый слой: 182 нейрона, функция активации ReLU.
2. Второй слой: 40 нейронов, функция активации ReLU.

3. Выходной слой: 5 нейронов, функция активации Softmax.

Перед скрытым слоем используется dropout с вероятностью отключения 0.1. В качестве оптимизатора использовался Adam. Функция потерь была выбрана categorical_crossentropy. Число эпох обучение 20 с размером batch равным 1000.

3. Заключение

В рамках данного исследования была продемонстрирована работоспособность предлагаемого подхода к классификации структурных элементов текста на основе графового представления из слов. Точность распознавания составила 0.793. Использование статистических характеристик построенного графа позволило добиться приемлемых результатов классификации. Дальнейшее исследование, направленное на применение графовых нейронных сетей для решения данной задачи, имеет потенциал для достижения результатов, сопоставимых с лидерами в этой области.

Благодарности

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 1023110300006-9).

Список литературы

- [1] S. Tsujimoto, H. Asada, Major components of a complete text reading system, Proceedings of the IEEE 80 (1992) 1133–1149. doi:10.1109/5.156475.
- [2] C. Da, C. Luo, Q. Zheng, C. Yao, Vision grid transformer for document layout analysis, in: ICCV, 2023.
- [3] Y. Huang, T. Lv, L. Cui, Y. Lu, F. Wei, Layoutlmv3: Pre-training for document ai with unified text and image masking, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022.
- [4] J. Wang, M. Krumdick, B. Tong, H. Halim, M. Sokolov, V. Barda, D. Vendryes, C. Tanner, A graphical approach to document layout analysis, in: International Conference on Document Analysis and Recognition, Springer, 2023, pp. 53–69.
- [5] R. Wang, Y. Fujii, A. C. Papat, General-purpose OCR paragraph identification by graph convolution networks, CoRR abs/2101.12741 (2021). URL: <https://arxiv.org/abs/2101.12741>. arXiv:2101.12741.
- [6] A. L. Maia, F. D. Julca-Aguilar, N. S. Hirata, A machine learning approach for graph-based page segmentation, in: 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2018, pp. 424–431. doi:10.1109/SIBGRAPI.2018.00061.
- [7] X. Zhong, J. Tang, A. J. Yepes, Publaynet: largest dataset ever for document layout analysis, in: 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2019, pp. 1015–1022. doi:10.1109/ICDAR.2019.00166.

- [8] R. W. Smith, An overview of the tesseract ocr engine, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) 2 (2007) 629–633. URL: <https://api.semanticscholar.org/CorpusID:7038773>.

Методы повышения точности пассивной локации импульсных источников геоакустических колебаний

Оксана Копылова ¹

¹ *Институт вычислительной математики и математической геофизики СО РАН, пр. акад. Лаврентьева, 6, 630090, Новосибирск, Россия*

Аннотация

Рассматривается проблема геофизического мониторинга импульсных источников, несущих геологические угрозы для окружающей социальной инфраструктуры. В качестве таких источников могут выступать карьерные и промышленные взрывы, падающие на землю осколки космических тел и др., являющиеся мощными источниками сейсмических и акустических волн – далее сейсмоакустических волн. Основными задачами здесь являются обнаружение, определение мощности и местоположения источников. Рассматриваются вопросы повышения точности и дальности обнаружения и локации импульсных источников с учетом параметров геометрии расстановки регистрирующих датчиков и измерения параметров сейсмоакустических волн. Предложенные теоретические подходы к решению задач реализованы в условиях натурного эксперимента.

Keywords

Импульсные колебания, сейсмоакустическая локация, поляризационный метод, полевой эксперимент.

1. Введение

Проблема геофизического мониторинга окружающей среды в мире связана с изучением экологической проблемы – влияния разного класса источников, порождающих мощные вибрации (сейсмические волны) в земле и акустические в атмосфере, на окружающую социальную среду и на человека. В качестве таких источников могут выступать карьерные и промышленные взрывы, падающие на землю тела при спутниковых запусках, осколки космических тел и др. [1-3]. Основными задачами здесь являются обнаружение, определение местоположения и мощности источников.

В решении задачи геолокации одним из основных является понятие фазового фронта и подобия волн в среде их распространения от источника. Подобие сигналов при этом оценивается функцией когерентности сигналов на выходах регистрирующих датчиков. В частности, на этих принципах базируется подход к определению направления на активные источники по их преобладающим спектральным составляющим шумов [4]. Другой подход к решению рассматриваемой задачи связан с теорией корреляционно-экстремальных систем [5], в основе которых лежит вычисление функции взаимной корреляции случайных процессов и определения координат главного экстремума этой функции, характеризующих положение лоцируемого источника. В работе [6] положение источника колебаний определяется по временам прихода поперечной и продольной волн. Способ гипербола, предложенный в [7] связан с получением разностей времен прихода Р-волны относительно опорного сейсмоприемника. В работе [6] для определения координат приводится вариант реализации метода гипербола с использованием данных от четырех пространственно разнесенных датчиков. Авторами патента [8] для повышения точности пеленгации объектов по данным сейсмических колебаний предлагается устройство с предварительной адаптивной фильтрацией

⁶ International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: oksana@opg.sccc.ru (A. 1)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.30

сигналов, состоящей из фильтров нижних частот и сглаживающего фильтра. Отфильтрованные сигналы поступают на вход многоканального корреляционного измерителя разности временных запаздываний.

В настоящей работе предлагаются подходы к решению задач повышения дальности обнаружения и точности локации импульсных источников с одновременным учетом сейсмических и акустических волн и принципа их подобия. Высокая эффективность предложенных подходов доказана результатами натуральных экспериментов.

2. Постановка задачи геолокации

С помощью линейки пространственно-распределенных векторно-фазовых датчиков в земле и атмосфере регистрируется волновое поле, представляемое в виде векторной функции:

$$\bar{A}(x, y, z, t) = \left\{ \bar{A}_k [x_k(t_i), y_k(t_i), z_k(t_i)] \right\}, \quad (1)$$

где $i=1..N$ – множество дискретных значений функций, $k=1..M$ – число датчиков в линейке $[x_k(t_i), y_k(t_i), z_k(t_i)]$.

В качестве начального на ориентацию датчиков накладываемся условие: компонента X ориентируется на север, Y – на восток, Z – вертикальная компонента.

Координатные функции как результат регистрации сейсмических колебаний в земле с учетом характеристик регистраторов описываются в виде:

$$x_k(t_i) = A_x(k)h_k L[s(t_i - \Delta t_k) + n_k(t_i)]. \quad (2)$$

Здесь A_x – амплитуда колебания по компоненте X на k -ом датчике, h_k – чувствительность датчика, L – оператор фильтрации сигнала, $s(t_i - \Delta t_k)$ – полезный сигнал, $n_k(t)$ – внешний шум с корреляционной функцией $r_k(\tau)$.

Решение задачи локации распадается на решение задач определения пеленга и на вычисления расстояния до него.

Дан вектор \bar{C} , определяющий расстановку сейсмических и акустических датчиков на плоскости. Необходимо определить времена вступлений сейсмических \bar{t}_s и акустических \bar{t}_a волн. На основе вычисленных времен определить координаты источников. Уравнение минимизации невязки имеет вид:

$$\Delta(x_c, y_c) = \left| \left\{ \bar{x}_{c_ист}, \bar{y}_{c_ист} \right\} - \left\{ \bar{x}_{c_изм}, \bar{y}_{c_изм} \right\} \right| \rightarrow \min, \quad (3)$$

где $\left\{ \bar{x}_{c_ист}, \bar{y}_{c_ист} \right\}$ – вектор истинных значений координат, $\left\{ \bar{x}_{c_изм}, \bar{y}_{c_изм} \right\}$ – вектор вычисленных значений координат.

3. Методика решения задачи

3.1. Предобработка данных и выделение информативных признаков

Одна из задач – обнаружение – связана с повышением соотношения «сигнал/шум», характеризующего точность алгоритмов обнаружения. Для этого необходима предварительная обработка данных на основе использования спектральных характеристик от заданных источников. Импульсные источники порождают в среде волны, модель которых описывается затухающей функцией времени, на которую, к тому же накладываются внешние шумы. Это определяет необходимость использования характеристик колебаний на основе оконного спектрального анализа Фурье для прослеживания динамики изменения спектра во времени.

Другой подход к выделению информативных признаков и уменьшению уровня фоновых шумов связан с использованием вейвлет-преобразования. Обработка сейсмических данных осуществляется с применением вейвлет-фильтрации, основанной на дискретном вейвлет-разложении сигнала по слоям детализации, что позволяет повысить отношение сигнал-шум в режиме обработки по одиночному каналу. Порядок и уровни вейвлет-коэффициентов

выбираются на основе анализа функции распределения энергии анализируемого сигнала по уровням разложения с учётом порядка вейвлета [9].

Обнаружение сигнала осуществляется на основе использования порогового энергетического критерия.

3.2. Локация с использованием линейки пространственно распределенных однокомпонентных датчиков

Локация с использованием линейки пространственно распределенных однокомпонентных датчиков при их линейной расстановке осуществляется на основе построения диаграммы коэффициента направленного действия (КНД). Зависимость КНД сейсмической антенны из n равномерно расставленных датчиков от временной задержки прихода волны между двумя соседними датчиками вычисляется согласно [10].

Вычисление коэффициента направленного действия производится по формуле:

$$К.Н.Д. = E_{сум} / E_{max} . \quad (4)$$

Рассматривается модификация алгоритма, основанная на взаимном перемножении соответствующий отсчетов сейсмограмм:

$$E_{сум} = \sum_n (f_1(t_n) * f_2(t_n + \Delta t) * \dots * f_k(f_n + \Delta t \cdot (k - 1)))^2 . \quad (5)$$

Пример расчета диаграммы КНД в случае использования одной линейки из 3 датчиков с шагом расстановки 105 метров на данных модельного эксперимента с использованием импульса Берлаге приведен на рис. 1 (а), в случае одной линейки из 5 датчиков с шагом 35 метров – на рис. 1 (б). Соотношение «сигнал/шум» задавалось равным 40, угол – 45 градусов, основная частота – 12 Гц. Из рисунков видно, что в первом случае точность определения направления на источник оказалась выше на 0,07°. Точность определения направления на источник в зависимости от шага расстановки и количества точек регистрации коррелирует с результатами анализа данных полевых исследований.

Совместный учет результатов вычислений согласно [10] и по формуле (5) при использовании двух взаимоперпендикулярных линеек пространственно разнесенных датчиков с общей серединной точкой позволяет определять направление на источник. Пример работы алгоритма для случая обработки данных от 5 датчиков, расположенных на взаимно перпендикулярных линейках по 3 датчика в каждой, представлен на рис. 1 (в, г). Датчики в линейках располагались на расстоянии около 103.5 метров друг от друга. Источник располагался на расстоянии около 800 метров от центра расстановки датчиков.

На рис. 1 (в) показан пример вычисленной диаграммы направленности по сейсмической волне, на рис. 1 (г) – по поверхностной акустической волне. Истинное значение угла составляет 269°. Видно, что на основе сейсмических колебаний точность определения направления на источник выше (отклонение составляет 1°), чем на основе акустических (отклонение составляет 4°), что объясняется дополнительным влиянием метеофакторов, в частности ветра, на распространение акустических колебаний.

3.3. Сейсмическая локация на основе трехкомпонентной векторной поляризации

В основе решения лежит поляризационный метод определения пеленга на источник [11]. В прямоугольной системе координат траектории колебаний (1) описываются полным вектором колебаний A с проекциями на горизонтальные и вертикальную составляющие A_x, A_y, A_z по осям x, y, z . Данные измерений представляются в виде набора точек, пространственное положение каждой из которых характеризуется радиус-вектором $\vec{A}_i \equiv (x_i, y_i, z_i)$. Ставится задача определения направления, задаваемого единичным вектором $\vec{p} \equiv (p_x, p_y, p_z)$ поляризации так, чтобы сумма квадратов расстояний всех точек от прямой, проведённой в этом направлении,

была минимальна. Для автоматизации процедуры решение предлагается искать с применением метода наименьших квадратов. Пример результата работы алгоритма в виде полярограммы приведен на рисунке 1 (д). Расстояние до источника составляло около 1 километра.

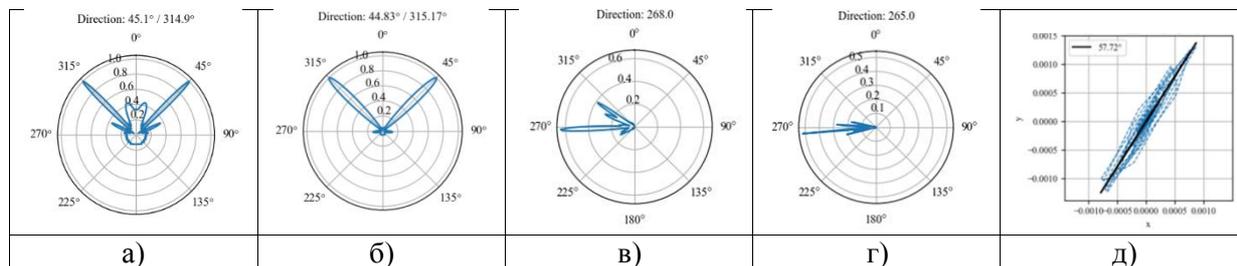


Рисунок 1: а) результат расчета КНД на модельных данных по одной линейке, состоящей из 3 датчиком с шагом 105 метров; б) результат расчета КНД на модельных данных по одной линейке, состоящей из 5 датчиком с шагом 35 метров; в, г) определение направления на источник по сейсмическим (в) и акустическим (г) колебаниям по двум линейкам. Шаг расстановки датчиков равен 103,5 метров; д) пример полярограммы

4. Заключение

В работе рассмотрена проблема, связанная с геоакустическим мониторингом разного класса импульсных источников - карьерных и полигонных взрывов, промышленных взрывов и др. Рассмотрена задача повышения дальности обнаружения и локации таких источников. Проведен сравнительный анализ точности работы алгоритма на данных сейсмических и акустических колебаний. Исследована зависимость точности работы алгоритма на модельных данных и данных реального эксперимента от количества используемых в линейке датчиков и от шага их расстановки. По результатам натурных экспериментов показана высокая эффективность предложенных подходов.

5. Благодарности

Работа выполнена в рамках госзадания FWNМ–2022–0004.

6. Литература

- [1] Thomas B Afeni, Stephen K Osasan. Assessment of noise and ground vibration induced during blasting operations in an open pit mine — A case study on Ewekoro limestone quarry, Nigeria. *Mining Science and Technology (China)*, Volume 19, Issue 4, 2009, Pages 420-424, ISSN 1674-5264. URL: [https://doi.org/10.1016/S1674-5264\(09\)60078-8](https://doi.org/10.1016/S1674-5264(09)60078-8).
- [2] Зыков В.С. Состояние вопроса по проблеме влияния промышленных взрывов при открытой разработке угольных месторождений на охраняемые объекты и окружающую среду. *Вестник Научного центра ВостНИИ по промышленной и экологической безопасности*, № 3, 2018. С. 51-56.
- [3] Пименов И.К., Дементьев Н.А., Кузнецова А.Д. Оценка шума от взрывных работ в карьере. // *Защита от повышенного шума и вибрации. Сборник трудов Всероссийской научно-практической конференции с международным участием. Под редакцией Н.И. Иванова. Санкт-Петербург, 2021. С. 87–97.*
- [4] Карлтон П.Н., Фидлер Р.У. Определение положения источника сейсмических колебаний методом пассивного эксперимента. В кн. «Анализ и выделение сейсмических сигналов». М.: «Мир». С.158-175.
- [5] Странгуль О.Н., Тарасенко В.П. Корреляционно-экстремальные системы навигации и локации подвижных объектов // *Автоматика и телемеханика*, № 7, 2001. С. 201–210.

- [6] Юлдашев Т. Модели распространения сейсмических волн в слоистой среде земного шара и алгоритмы определения параметров источника / Т. Юлдашев; отв. Ред. О.М. Набиев; АН Респ. Узбекистан, НИИ «Алгоритм-Инжинеринг» . – Т.: Фан, 2011. 232 с. ISBN 978-9943-19-113-6
- [7] Абуталиев, Ф.Б. Определение координат близких землетрясений на ЭВМ М-20 с последующей оценкой их точности // Ф.Б. Абуталиев, Е.М. Бутовская, Л.Н. Плотникова и др.; Применение ЭВМ к решению задач математической физики. –Ташкент, 1967. 118с.
- [8] Патент на изобретение 2536087 С1 Российская Федерация. Устройство для адаптивной сейсмической пеленгации объектов. МПК G08B 13/181, G01V 1/16. Дудкин В.А., Вольсков А.А., Иванцов Д.С., Прокина Н.В. Заявка № 2009119839/08 от 25.05.2009. Опубл. 27.01.2011 г.
- [9] Хайретдинов М.С., Авроров С.А. Обнаружение и распознавание взрывных источников. // Вестник НЯЦ РК, №12, 2012. С.17–24.
- [10] Хайретдинов М.С., Клименко С.М. Программная система автоматизированной локации и визуализации сейсмических источников. // Вестник НЯЦ РК, Вып. 2(18), 2004. С. 70-76.
- [11] Гальперин Е.И. Поляризационный метод сейсмических исследований. М., «Недра», 1977. 277 с.

Защита от атак с использованием омоглифов на модели NLP

Кузнецов А.А.^{1,2}, Михайлов А.А.^{1,2}

¹Институт динамики систем и теории управления им. В.М. Матросова Сибирского отделения Российской академии наук, 664033, Иркутск, ул. Лермонтова, д. 134, Российская Федерация

²Институт системного программирования им. В.П. Иванникова Российской академии наук, 109004, г. Москва, ул. Александра Солженицына, д. 25, Российская Федерация

Аннотация

В статье рассматривается проблема атак на системы обработки естественного языка (NLP) с использованием визуально похожих символов — омоглифов. Омоглифы представляют собой графически идентичные или похожие символы, имеющие разное значение. В статье приводится анализ существующих методов защиты от таких атак и предлагается новый подход на основе модели T5, обученной под корректировку орфографических ошибок на английском языке.

Ключевые слова

Обработка естественного языка, защита от атак на языковые модели, омоглифы,

1. Введение

С развитием глубоких нейронных сетей в области обработки естественного языка [1] стало ясно, что подобные системы уязвимы для атак, искажающих входные данные.

Раньше атаки на языковые модели были заметными для человеческого глаза, например, орфографические ошибки или перефразирования. Сейчас же набирают популярность атаки с использованием омоглифов — графически одинаковых или похожих друг на друга знаков, имеющих разное значение. Они незаметны для человека, но ощутимо различаются при обработке.

Возможность изменить текст, не изменяя его визуально, может быть использована многими злоумышленниками для обхода механизмов фильтрации контента, некорректного машинного перевода, ухудшения качества запросов и индексации поисковых систем. Согласно статье [2] многие системы обработки естественного языка не обращают внимания на возможные проблемы: от снижения производительности до некорректной работы.

Понимание сущности атак с использованием омоглифов и их потенциальных последствий для безопасности систем станет ключом для разработки эффективных методов

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

✉ viirtualkuz@yandex.ru (А.А.); mikhailov@icc.ru (А.А.)

🆔 0009-0001-1105-0076 (А.А.); 0000-0003-4057-4511 (А.А.)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings (icc-de.icc.ru)

DOI: 10.47350/ICCS-DE.2024.31

защиты. Анализ реальных случаев атак с использованием омоглифов поможет выявить наиболее уязвимые области и разработать соответствующие меры противодействия.

2. Предлагаемый подход

В статье предлагается новый подход защиты от атак с использованием омоглифов на основе модели T5. Модель T5 — text to text transfer transformer [3, 4], представляет собой предобученную командой SberDevices [5] модель, которая способна корректировать орфографические ошибки на английском языке. Для генерации обучающего корпуса существует два разных алгоритма: с использованием случайной замены символа на его омоглиф и на основе генетического алгоритма [6]. Составленный словарь омоглифов рассчитан на замену латинских символов. В среднем по 6 омоглифов на каждый символ. Размер обучающего корпуса превышает 400 тыс. элементов (алгоритм случайной замены).

Промежуточный результат после первых попыток обучения составляет свыше 80% (Ассигасу). Полученные результаты демонстрируют перспективность выбранного подхода и эффективность использования модели T5 для решения задачи защиты от атак с использованием омоглифов. Однако, несмотря на положительные результаты, необходимо дальнейшее развитие проекта, включающее в себя работу с русским языком (кириллицей), расширение существующего словаря омоглифов, генерацию большого обучающего корпуса с использованием генетического алгоритма, добавление новых метрик в процесс оценивания работы модели и оптимизацию модели.

3. Заключение

Таким образом, предложенный подход на основе модели T5 является перспективным направлением для разработки эффективных методов защиты от атак с использованием омоглифов. Дальнейшее развитие проекта позволит расширить область применения предложенного метода и повысить уровень безопасности систем обработки естественного языка.

Благодарности

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 1023110300006-9).

Список литературы

- [1] J. O'Connor, I. McDermott, NLP, Thorsons, 2001.
- [2] N. Boucher, I. Shumailov, R. Anderson, N. Papernot, Bad characters: Imperceptible nlp attacks, in: 2022 IEEE Symposium on Security and Privacy (SP), IEEE, 2022, pp. 1987–2004.

- [3] A. Gillioz, J. Casas, E. Mugellini, O. Abou Khaled, Overview of the transformer-based models for nlp tasks., In 2020 IEEE 15th Conference on Computer Science and Information Systems (FedCSIS) (2022) 179–183.
- [4] M. Guo, J. Ainslie, D. Uthus, S. Ontanon, J. Ni, Y. H. Sung, Y. Yang, Longt5: Efficient text-to-text transformer for long sequences., 2020. URL: <https://doi.org/10.48550/arXiv.2010.11934>.
- [5] N. Martynov, M. Baushenko, A. Kozlova, K. Kolomeytseva, A. Abramov, A. Fenogenova, A methodology for generative spelling correction via natural spelling errors emulation across multiple domains and languages, in: Y. Graham, M. Purver (Eds.), Findings of the Association for Computational Linguistics: EACL 2024, Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 138–155. URL: <https://aclanthology.org/2024.findings-eacl.10>.
- [6] L. Haldurai, T. Madhubala, R. Rajalakshmi, A study on genetic algorithm and its applications, International Journal of computer sciences and Engineering 4 (2016) 139.

Сравнение моделей Игрового Искусственного Интеллекта

Алексей Линявский¹

¹Иркутский государственный университет путей сообщения, ул. Чернышевского, д. 15, Иркутск, 664074, Россия

Аннотация

Рассматриваются существующие модели игрового искусственного интеллекта. Приведено краткое описание базовой реализации и работы моделей игрового искусственного интеллекта Finite State Machine (FSM), Behavior Tree, Utility AI, Goal-Oriented Action Panner (GOAP), Hierarchical Task Network (HTN). Для каждой модели выделены положительный и отрицательные стороны разработки, дальнейшего поддержания и использования. Произведён анализ и сравнение со стороны рационального агента на основе описания базовой реализации каждой модели.

Ключевые слова

Игровой Искусственный Интеллект, искусственный интеллект, сравнение, GOAP, Behavior Tree, HTN, Utility AI, FSM.

1. Введение

При выборе искусственного интеллекта для управления игровым персонажем или автономными роботами существует проблема выбора искусственного интеллекта, который будет принимать решения о дальнейших действиях агента.

В данной статье представлено сравнение большинства популярных моделей игрового искусственного интеллекта.

2. Модели

Сравнение будет основано на следующих моделях игрового искусственного интеллекта:

1. FSM - Аппарат конечных состояний;
2. Behavior Tree – Дерево поведения;
3. Utility AI – Утилитарный искусственный интеллект;
4. GOAP – Ориентированный на задаче планировщик действий;
5. HTN – Сеть иерархических задач.

Каждая модель, описанная ниже, представлена в своей "минимально комплектации", поскольку для большинства реализаций эти модели немного модифицируются, и сравнение такого количества модификаций не представляется возможным.

2.1. FSM

Finite State Machine (FSM) [1, 3] считается одним из самых простых реализаций игрового искусственного интеллекта.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: inetlsec@mail.ru;

ORCID: 0009-0008-0412-9158



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.32

Каждое состояние FSM описывает действие искусственного интеллекта и имеет возможность перехода в другое состояние (или состояния).

Для перехода между состояниями должно быть соблюдено условие перехода.

FSM обеспечивает достаточную гибкость разработки и использования при малых затратах ресурсов. Однако расширение подобной модели искусственного интеллекта невозможно без прямого взаимодействия с кодом FSM. Это, в свою очередь, влечёт за собой увеличение затрат на разработку подобной модели, а расширение большого FSM может ещё и значительно замедлить работу над моделью.

2.2. Behavior Tree

BT (Behavior Tree) [1, 3, 8] представляет собой дерево, которое состоит из root (корня), node (технических вершин) и leaf (вершин-листьев).

Root выполняет роль бесконечного цикла в дереве, постоянно перебирая его детей-вершин в зависимости от ответа, который они вернут. node содержат в себе описание возвращаемого ответа и порядок перебора детей-вершин. Обязательно содержит хотя бы 1 вершину в детях.

Leaf описывают непосредственно действие, которое необходимо выполнить, а также decorator - описание выполнения действия.

Leaf может вернуть 3 состояния: Success (Успех), Failed (неудача) и Running (в процессе). Эти состояния обрабатываются decorator, после чего он их передаёт в node. node, в зависимости от полученного ответа, переходит либо к следующему своему ребёнку, либо возвращает ответ своему родителю. И так, пока ответ не достигнет root.

Дерево поведения можно представить как "памятку при пожаре", которой следует агент искусственного интеллекта. Это позволяет достаточно точно описать все необходимые действия, которые искусственный интеллект должен предпринять.

Деревья поведения требуют от агента искусственного интеллекта знания об окружающей среде, что также требует и ориентирование искусственного интеллекта. Также необходим объектно-ориентированный подход при создании вершин, а значит для Деревьев поведения необходим высокоуровневый язык программирования.

2.3. Utility AI

Концепция Utility AI (UAI) [5, 6] достаточно проста:

Агент искусственного интеллекта знает о каких-то возможных ему действиях, сравнивает релевантность (вес) этих действий, находит наиболее ему подходящее и выполняет это действие.

Релевантность может быть представлена нелинейной зависимостью от нескольких переменных, что позволяет гибко настраивать "ценность" действий в глазах искусственного интеллекта.

UAI обеспечивает гибкое поведение искусственного интеллекта в зависимости от окружающей среды, а расширение возможностей модели UAI не займёт больших затрат ресурсов.

Однако UAI необходимо знать все доступные ему действия и сравнивать их вес между собой. Иначе говоря, оптимизация поиска оптимального действия при их большом количестве может быть затруднительна. А поскольку вес изменяется каждую условную единицу времени, то UAI может показать неэффективное использование ресурсов на достаточно большом количестве действий. Также, UAI должен знать об окружающей среде, в которой он находится, иначе говоря, он должен как-то в ней ориентироваться.

2.4. GOAP

Goal-Oriented Action Planner (GOAP) [1, 2] строит цепочку выполняемых действий исходя из текущей задачи агента искусственного интеллекта.

Задача – некоторое конечное состояние, которого желает достичь агент искусственного интеллекта. Каждая задача имеет несколько параметров, на основе которых и она и выбирается планировщиком задач: Начальное состояние агента, Конечное состояние агента, Вес.

Начальное состояние – состояние, в котором необходимо находиться агенту искусственного интеллекта.

Конечное состояние – состояние, в которое перейдет агент искусственного интеллекта в случае успеха выполнения задачи.

Действие – непосредственно действие, которое выполнит агент искусственного интеллекта.

Каждое действие состоит из нескольких параметров: Непосредственно действие, Конечное состояние, Начальное состояние, Условие выполнения и Вес.

GOAP хоть и прост в своей концепции, требует работу с объектно-ориентированной парадигмой, поскольку реализация действий и задач без ООП представляется затруднительной. Иначе говоря, GOAP, в большинстве случаев, может работать только на высокоуровневых языках.

GOAP требует от агента искусственного интеллекта знание о мире, а потому ему необходимо какое-то ориентирование в окружающей среде.

Однако, взамен на вышеперечисленные требования, GOAP обеспечивает достаточно высокую интеллектуальность искусственного интеллекта, а также очень просто экстенсивное расширение возможностей искусственного интеллекта.

И, что тоже важно, GOAP может быть применён не к одному агенту искусственного интеллекта, а сразу к нескольким. Это позволяет построить на GOAP роевой интеллект.

2.5. HTN

Hierarchical Task Network (HTN) [4, 7] состоит из плана, который выполняет агент искусственного интеллекта. План в себе содержит 2 типа задач: Примитивные и Составные. Примитивные задачи, как правило, олицетворяют собой конкретное действие, которое выполняет агент искусственного интеллекта.

Составные задачи включает в себя как примитивные, так и другие составные задачи.

Образованный план выглядит как иерархическое представление папок в операционных системах, за что и получил своё название.

HTN позволяет искусственному интеллекту "мыслить" более абстрактно, так, как это делает человек. Это позволяет искусственному интеллекту создавать достаточно сложные и длинные цепочки действий. А гибкость добавления задач может стать сильным подспорьем в выборе именно данной модели.

В то же время, HTN, также как и GOAP, требует ориентацию в окружающем пространстве и ООП. А разработка HTN потребует хорошего специалиста, который понимает, как работает эта модель.

3. Выводы

Каждая модель Игрового искусственного интеллекта уникальна в своём исполнении. Иначе говоря, представленные выше модели представляют собой некоторую идею (концепцию), которую эти модели отражают. Как правило, перед реальным использованием этих моделей их немного модифицируют под конкретную задачу, по аналогии с Очередью (Queue) и Кучей (Stack). Так что сравнение моделей ограничено их базовыми вариациями, поскольку учесть все модификации моделей не представляется возможным.

Все сравнения и возможности применения построены с точки зрения рационального агента, который знает обо всех остальных возможных моделях.

FSM, несмотря на свою гибкость, применим только для простых систем, поскольку обеспечить его поддержку в дальнейшем может быть трудозатратно из-за огромной диаграммы состояний.

BT обеспечивает точность действий агента искусственного интеллекта, однако недостаточно гибок, в сравнении с другими моделями. Остаётся проблема расширения искусственного интеллекта. Требует ориентирования в пространстве для своей работы.

Гибкость UAI позволяет применить его в совершенно разной среде, однако на большом количестве необходимых действий может значительно замедлить свою работу. Также требует ориентирования в пространстве.

GOAP позволяет создать гибкий искусственный интеллект, который позволит агенту исполнять сложные действия. Однако потребует ориентирования в пространстве для своей работы, а также высокоуровневых языков (напр. C++, Java, C# и др.).

HTN обеспечивает весьма гибкую и легко расширяемую систему задач искусственного интеллекта, но требует хорошего специалиста для создания такой модели. Также, как и с GOAP, требует высокоуровневый язык и ориентирование в пространстве.

4. Список литературы

- [1] DigDigger, How does videogame AI make its decisions? (FSM, Behaviour Trees, BDI, GOAP) | Bitwise. URL: https://www.youtube.com/watch?v=ValJk15l_y8
- [2] sgkryvenko, Проектирование непредсказуемого интеллекта в играх. Часть 2 интеллект толпы. URL: <https://habr.com/ru/articles/654601/>
- [3] Kylotan, The Total Beginner's Guide to Game AI. URL: <https://www.gamedev.net/tutorials/programming/artificial-intelligence/the-total-beginners-guide-to-game-ai-r4942/>
- [4] T. Humphreys, Exploring HTN Planners through Example. URL: https://www.gameai.pro/GameAIPro/GameAIPro_Chapter12_Exploring_HTN_Planners_through_Example.pdf
- [5] This is Vini, Utility AI with examples. URL: <https://www.youtube.com/watch?v=d63hbJYYqM8>
- [6] AI and Games, How Utility AI Helps NPCs Decide What To Do Next | AI 101. URL: <https://www.youtube.com/watch?v=p3Jbp2cZg3Q>
- [7] AI and Games, HTN Planning in Transformers: Fall of Cybertron | AI and Games #14. URL: <https://www.youtube.com/watch?v=kXm467TFTcY>
- [8] Cyberstars Entertainment, Урок 6 | Unreal Engine 4 Искусственный интеллект - Дерево поведения ИИ | AI Behavior tree Blackboard. URL: https://www.youtube.com/watch?v=hWfIRS_DSuE

Среда программирования киберфизических систем в парадигме машин состояний

Михаил Андреевич Чекан¹

¹ *Институт динамики систем и теории управления им. В.М. Матросова Сибирского отделения Российской академии наук, Россия, 664033, г. Иркутск, ул. Лермонтова, 134*

Аннотация

Доклад посвящён среде визуального программирования, разработанной для национальной киберфизической платформы. Среда реализует визуальную парадигму на основе иерархических машин состояний и пиктографического языка, и тем самым служит связующим звеном между образовательными играми платформы «Берлога» и проектированием настоящих систем на основе микроэлектроники и программирования. Доклад описывает текущее состояние среды и перспективы дальнейшей разработки.

Ключевые слова

Автоматное программирование, среды разработки, киберфизические системы

1. Мотивация

Киберфизическая система (КФС), согласно [1], – это комплексная система физических элементов и их цифровых двойников в вычислительном слое управления, которая постоянно получает данные из окружающей их среды и использует их для оптимизации процессов достижения установленных целей. КФС становятся всё более актуальными в общем поле, являясь по сути закономерным способом представления сложных инфраструктурных систем, например, в энергетике, на транспорте, в сельском хозяйстве и других важных сферах человеческой деятельности. При этом исследования и академическое рассмотрение КФС ведётся достаточно давно [2].

Национальная киберфизическая платформа (НКФП) «Киберриада» была создана в 2022 году как проект по привлечению инженерных кадров и созданию новых инструментов для проектирования микроэлектронных и киберфизических систем [3]. Здесь следует выделить два направления деятельности.

Первое – популяризация технического творчества и технологического образования в сфере микроэлектроники и киберфизики, создание траектории от школьной программы и технологических кружков до профильных курсов и решения промышленных задач. Это направление определило первое решение – создание платформы «Берлога» [4], набора образовательных игр в едином сеттинге. В настоящее время доступны игра «Защита пасеки» в жанре Tower Defense и ролевая игра-квест «Академия». В этих играх присутствует программирование как игровая механика, овладение которой необходимо для успешного прохождения игровых миссий. Важной особенностью «Берлоги» является двусторонняя связь с реальной образовательной сферой через события и достижения в цифровом портфолио школьника (или студента) на платформе «Талант» Кружкового движения Национальной технологической инициативы (НТИ).

Второе направление – разработка государственных стандартов для КФС и средств их проектирования. Кружковое движение совместно со специалистами формирует стандарт НКФП,

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: chekoopa@mail.ru (М.А. Чекан)

ORCID: 0000-0001-7622-421X (М.А. Чекан)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.33

а также инструментальную базу для проектирования КФС. Задача стандартов и инструментария – поднять качество создаваемых систем и доступность их создания более широкому кругу пользователей.

Одним из решений стало принятие парадигмы программирования расширенных иерархических машин (ПРИМС) как основной при проектировании КФС. ПРИМС опирается на практическое применение стандарта UML Statecharts [5] и родственна автоматному программированию [6]. Парадигма ПРИМС реализована в играх «Берлоги», но для перехода в разработку реальных систем необходима среда разработки, выступающая связующим звеном. Им стала Lapki IDE, графическая среда программирования в парадигме ПРИМС, совместимая как с играми «Берлоги», так и аппаратными платформами. Иными словами, данный доклад посвящён состоянию разработки Lapki IDE, текущим выводам и направлениям дальнейшей разработки и исследований. Доклад является закономерным продолжением доклада на ICCS-DE 2023 [7].

2. Текущий статус

На момент написания тезисов среда Lapki IDE находится в активной разработке силами студенческой команды под руководством автора и размещена на платформе GitHub под эгидой Кружкового движения [8]. Существует публичная бета-версия, в полной мере реализующая запланированную в рамках первой итерации функциональность, а именно разработку схем для игры «Защита пасеки» на платформе «Берлога», а также для платы Arduino Uno. Архитектура среды в большинстве аспектов соответствует запланированной (рисунок 1).



Рисунок 1: Схема-план архитектуры Lapki IDE

Так, основным компонентом является клиентская часть на основе веб-фреймворка Electron, в которой реализован редактор ПРИМС с доступом к просмотру промежуточного кода. Клиент взаимодействует с модулем загрузчика, который может запускаться вместе с клиентом или размещаться в облаке (для удалённой прошивки). Модуль компиляции, транслирующий схему ПРИМС в формат целевой платформы (игра «Берлоги» или файл прошивки), размещается на отдельном сервере. Также реализована система документации, в настоящий момент – на основе веб-сервера с HTML-файлами.

Текущая версия IDE уже демонстрирует удобство визуальной парадигмы для быстрой разработки базовых устройств на основе Arduino, а также простоту внесения новой функциональности в существующие схемы. Заявленные ранее принципы проекта «Пиктограмма» уже проявляются в той или иной мере, будь то «открытый капот» (доступность промежуточных артефактов разработки), «разработка в состоянии потока» (продолжающиеся

эксперименты по улучшению пользовательского опыта и интуитивности создания РИМС), «модульность и транс-облачность» (максимальное вынесение модулей в облако)

3. Перспективы

Рисунок 1 также отражает дальнейшие направления разработки. В первую очередь, это развитие поддержки STM32 как платформы, на которой планируется создание ряда крупных образовательных проектов в рамках НКФП. Для повышения качества разработки и отладки предлагается внедрение в загрузчик монитора последовательного порта. Рассматриваются гипотезы внедрения в клиент симулятора одной из существующих платформ или специальной тренировочной платформы с визуальным симулятором, позволяющим ознакомиться с ПРИМС без дополнительного оборудования. Кроме этого, исследуется возможность разработки комплекса устройств с пересекающейся функциональностью как одного документа-проекта с несколькими машинами состояний и описанием схемотехники каждого устройства.

Для интеграции с другими проектами НКФП предложен формат Cyberiada-GraphML как совместимое расширение существующего формата представления графов, предоставляющее необходимые атрибуты для описания РИМС, максимально соответствующего разрабатываемому Кружковым движением стандарту парадигмы. Библиотека для чтения формата активно разрабатывается для языков C, C++, TypeScript и Python, также его поддержка реализуется в C# для игр платформы «Берлога» (тем самым планируется переход к единому с IDE формату).

4. Заключение

В докладе рассмотрен текущий статус разработки среды визуального программирования, создаваемой в рамках национальной киберфизической платформы под руководством автора. Существует публичная версия Larpi IDE, совместимая с «Защитой пасеки», и способная компилировать (посредством облачного модуля) и прошивать программы под аппаратные платформы Arduino Uno и Arduino Micro. Разработка среды активно продолжается, в ряде аспектов совместно с Кружковым движением и командой разработчиков игр «Берлоги».

5. Список литературы

- [1] Е.И. Громаков, А.А. Сидорова, Современные технологии. Киберфизические системы, Изд-во Томского политехнического университета, г. Томск, 2021.
- [2] Е.Н. Мещерякова, От сложных технических систем к киберфизическим системам. Современные информационные технологии, № 30, 2019, с. 51-54.
- [3] Кружковое движение, Владимир Путин поддержал запуск Национальной киберфизической платформы «Восток», 21 июля 2022. URL: <https://team.kruzhok.org/news/post/vladimir-putin-podderzhal>.
- [4] Национальная киберфизическая платформа, 2024. URL: <https://platform.kruzhok.org>.
- [5] M. Samek, Practical UML statecharts in C/C++: event-driven programming for embedded systems, 2nd. ed., Elsevier Inc., Oxford, UK, 2009.
- [6] Н. И. Поликарпова, А. А. Шалыто, Автоматное программирование, Издательство «Питер», г. Санкт-Петербург, 2009.
- [7] М. Чекан, Разработка среды пиктографического программирования киберфизических систем в парадигме машин состояний, Материалы V Международного семинара по информационным, вычислительным и управляющим системам для распределенных сред (ICCS-DE 2023), ИДСТУ СО РАН, г. Иркутск, 2023 года, с. 149-151.
- [8] Организация «Кружковое движение» на платформе GitHub. URL: <http://github.com/kruzhok-team>.
- [9] Образовательная игра «Орбита», 2023. URL: <http://orbitagame.ru>

Средство мониторинга узлов системы управления сервис-ориентированными научными приложениями

Дмитрий Яковлев¹, Михаил Воскобойников¹, Роман Костромин¹

¹Институт динамики систем и теории управления им. В.М. Матросова СО РАН, ул. Лермонтова, д. 134, 664033, Иркутск, Россия

Аннотация

В работе рассматривается архитектура программного средства, предназначенного для мониторинга (сбора, хранения и агрегирования ключевых метрик) производительности вычислительных узлов в гетерогенной распределенной вычислительной среде с целью предоставления актуальных метрик системе управления сервис-ориентированными научными приложениями. Полученные сведения позволяют снизить неопределенность при планировании вычислений, улучшить загрузку вычислительных узлов, а также сократить время проведения вычислительного эксперимента.

Ключевые слова

Гетерогенная распределенная вычислительная среда, вычислительные процессы, мониторинг, управление вычислениями

1. Введение

Система мониторинга является неотъемлемой частью управления распределенными вычислениями [1]. Ее основная задача — предоставление актуальных и достоверных сведений о состоянии всех компонентов гетерогенной распределенной вычислительной среды. Она включает в себя сбор и анализ данных о загрузке процессоров, использовании памяти, состоянии сети, доступности и состоянии хранилищ данных, а также о текущем состоянии выполняемых задач. Существующие системы мониторинга, такие как NetData, Nagios, Zabbix и др. [2], ориентированы в основном на администраторов ресурсов, а предоставляемые ими сведения порой являются избыточными, т.к. как необходимо собирать ограниченный набор метрик при планировании вычислительных экспериментов в задаче управления сервис-ориентированными научными приложениями. Зачастую такие приложения и сам вычислительный процесс представлены в виде workflow (научный рабочий процесс). Во время вычислений используются разнородные вычислительные ресурсы, в том числе облачные. Поэтому актуализируется разработка специализированных программных средств, которые позволят собирать необходимые метрики из разных источников (типов ресурсов) и предоставлять их системе управления вычислениями в нужном виде [3].

2. Архитектура программного средства

Архитектура программного средства представлена на рис. 1. На схеме изображены три узла («Node 1», «Node 2» и «Node 3» соответственно), с развернутыми на них модулями программного средства. Узел под названием «Node 2» представляет собой основной узел.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia

EMAIL: dimasgamer353@gmail.com (A. 1); mikev1988@mail.ru (A. 2); roman@kostromin.net (A. 3)

ORCID: 0009-0009-7973-1685 (A. 1); 0000-0003-3034-4907 (A. 2); 0000-0001-8406-8106 (A. 3)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2024.34

Сервис под названием «Exporter0», обращается ко всем узлам из списка, для получения информации о их доступности. Prometheus обращается к экспортерам на всех узлах через их API для сбора метрик и сохранения их в базу данных. Все это происходит с помощью pull запросов Prometheus к экспортерам. Сервис под названием «Backend» обращается к Prometheus через API, чтобы получать данные из базы данных и предоставлять эти же данные после обработки через собственное API.

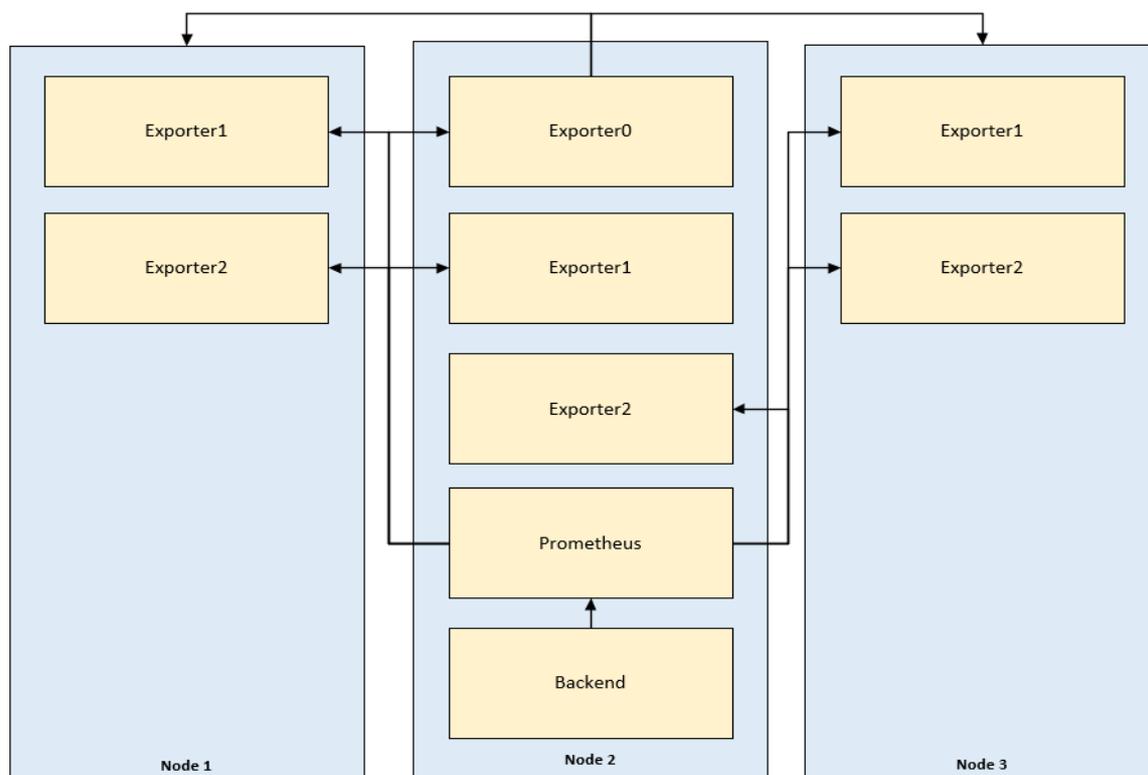


Рисунок 1: Архитектура программного комплекса мониторинга

Программное средство состоит из трех модулей (модуль управления, модуль сбора характеристик узла, модуль мониторинга процессов). В состав каждого модуля входят сервисы, которые были реализованы на языке программирования Python с использованием библиотек FastAPI и prometheus_client, которая является клиентской библиотекой для создания экспортера на языке программирования Python. Рассмотрим каждый модуль подробнее.

Модуль управления. Состоит из трех сервисов: Prometheus, Backend и Exporter0. Является основным модулем программного комплекса. Данный модуль должен быть развернут на основном узле и работать на постоянной основе, поскольку он содержит базу данных и сервисы, которые должны работать постоянно.

Модуль сбора характеристик узла. Состоит из одного сервиса: Exporter1. Данный модуль должен быть на каждом узле, с которого необходимо собирать характеристики о компонентах устройства.

Модуль мониторинга процессов. Состоит из одного сервиса: Exporter2. Данный модуль должен быть развернут на каждом узле, на котором необходимо выполнять мониторинг процессов. Модули сбора характеристик узла и мониторинга процессов не должны работать на постоянной основе и запускаются по требованию.

3. Заключение

Программное средство было развернуто и протестировано в испытательном стенде для мониторинга работы системы, реализующей методы структурно-параметрической оптимизации

модели энергетического комплекса. Целью выполнения мониторинга являлось определение вычислительных характеристик работы системы для определения наилучшего метода структурно-параметрической оптимизации. В качестве измеряемых параметров работы системы были выбраны расход оперативной памяти и использование процессора. Полученные сведения позволили учитывать при планировании критерии выполнения задачи и доступные квоты при распределении вычислительных модулей между доступными вычислительными ресурсами гетерогенной среды.

4. Благодарности

Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации, проект № FWEW-2021-0005 «Технологии разработки и анализа предметно-ориентированных интеллектуальных систем группового управления в недетерминированных распределенных средах» (рег. № 121032400051-9).

5. Список литературы

1. Костромин Р.О. Особенности реализации системы мониторинга гетерогенной вычислительной среды // Современные наукоемкие технологии. 2023. № 12. Ч. 3. С. 86–94.
2. Stefanov K.S., Pawar S., Ranjan A., Wandhekar S., Voevodin V.V. A Review of Supercomputer Performance Monitoring Systems // Supercomputing Frontiers and Innovations. 2021. Vol. 8. Is. 3. P. 62-81. DOI: 10.14529/jsfi210304.
3. Черных А.Н., Бычков И.В., Феоктистов А.Г., Горский С.А., Сидоров И.А., Костромин Р.О., Еделев А.В., Зоркальцев В.И., Аветисян А.И. Смягчение неопределенности при разработке научных приложений в интегрированной среде // Труды Института системного программирования РАН. 2021. Т. 33. № 1. С. 151-172. DOI: 10.15514/ISPRAS-2021-33(1)-11.

Научно-организационный отдел
Федерального государственного бюджетного учреждения науки
Института динамики систем и теории управления имени В.М. Матросова
Сибирского отделения Российской академии наук
664033, Иркутск, ул. Лермонтова, д. 134
E-mail: rio@icc.ru

Подписано к печати 12.07.2024 г.
Формат бумаги 60×84 1/16, объем 13,88 п.л.
Заказ 1. Тираж 1 экз.

Отпечатано в ИДСТУ СО РАН