Experience of Using Content Patterns in the Development of Ontologies of Scientific Subject Areas

Yu. A. Zagorulko[a],* (ORCID: 0000-0002-7111-6524) and **G. B. Zagorulko**[a],** (ORCID: 0000-0003-2155-1357)

[a] Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090 Russian Federation
*e-mail: zagor@iis.nsk.su
**e-mail: gal@iis.nsk.su

Abstract—At present, ontology is widely used for the formalization and systematization of knowledge and data in various subject areas. To facilitate and accelerate the process of developing ontologies, ontology design patterns have recently been used, which are formal descriptions of solutions to typical problems of ontological modeling proven in practice. The most popular type of ontology design patterns are content patterns that support a uniform and consistent representation of all the entities of the simulated subject area. The article gives a brief overview of ontology design patterns and proposes a methodology for constructing ontologies of scientific subject areas, the main feature of which is the use of basic ontologies and content patterns, and analyzes the experience of using this methodology in the construction of ontologies of various scientific subject areas.

INTRODUCTION

Owing to the fact that ontology has proved to be a convenient and effective means of formalizing and systematizing knowledge and data in various subject areas [13], there is an urgent need for effective methods and tools for constructing ontologies. This need is especially acute when constructing ontologies of scientific subject areas (SSAs), which usually include subject areas covering a certain scientific discipline or field of scientific knowledge in all its aspects [4].

Modern methods of development of ontologies [19] can be divided into three groups: (1) methods that ensure the development of ontology from scratch; (2) methods of creating ontologies from ready-made blocks; (3) methods of automatic construction of ontologies.

Methods of developing ontology from scratch [5, 7, 14, 23] are the most time-consuming and complex. They require the involvement of experienced specialists in the field of ontological engineering.

The second group includes less time-consuming methods [9, 22, 24], using previously created basic ontology and/or fragments (patterns) thereof, which can be specialized for a specific subject area. The advantage of these methods is that they make it possible to involve in the process of building and maintaining ontologies experts in subject areas, which is one of the most important modern requirements for the process of developing knowledge-based systems [11].

Methods of automatic construction of ontologies on the basis of thematic body of texts [2] or web resources [21] seem to be the most promising and least labor-intensive, but when using them in pure form, it is not possible to build SSA ontologies of good quality [8]. In this regard, in practice, such methods are often used to automatically populate ontologies built by "manual" methods [8, 28].

In this article, we will consider the method of constructing ontologies oriented to knowledge engineers and subject experts and based on the use of ontology design patterns (ODPs) [3, 9, 26], which are formal descriptions of solutions to typical problems of ontological modeling proven in practice. In particular, fragments of basic ontologies presented in the OWL language can act as patterns [1]. Thus, this method belongs to the second group.

The article gives a brief overview of ODPs, analyzes the problems of their application, and describes the methodology proposed by the authors of the construction of SSA ontologies, which uses basic ontologies and ODPs, as well as the experience of its application in the development of ontologies for various SSAs.

1. ONTOLOGY DESIGN PATTERNS AND PROBLEMS OF THEIR APPLICATION

As mentioned above, ODPs serve to describe solutions to problems often encountered in the development of ontologies [9]. Depending on the type of problems that ODPs are designed to solve, structural patterns, correspondence patterns, content patterns, reasoning patterns, presentation patterns, and lexico-syntactic patterns are distinguished. (Such a typology of patterns was proposed in the framework of the NeOn project [15].)

In the development of ontologies, knowledge engineers and subject matter experts mainly use structural patterns, content patterns, and presentation patterns.

Structural patterns either fix ways to solve problems caused by limitations of expressive capabilities of ontology description languages or set a common (modular) structure and type of ontology. Patterns of the first type are called logical patterns; patterns of the second type are called architectural patterns.

Content patterns are descriptions of typical fragments of ontologies, on the basis of which ontology of various subject areas can be built.

Presentation patterns determine the recommendations for naming and annotating the elements of ontology, the use of which should increase the readability of ontology, as well as the convenience and ease of use of it.

At the moment, there are several catalogs of ODPs [17, 18] available on the Internet. The most representative of them is posted on the portal of the Association for Ontology Design and Patterns [17], created as part of the NeOn project [15].

Despite the fact that the use of ODPs can save human resources and improve the quality of the developed ontologies, they have not yet found wide practical application because of a number of problems arising from their use.

The first problem is related to the difficulties of reusing the ODPs developed by other developers—often it is difficult for the developer of a new ontology to understand the semantics that its authors laid in a particular pattern.

The second problem is caused by the lack of convenient tools for the development of ontologies that support the use of ODPs. Here one can note plugins for the tool for the development of ontologies of the NeOn project [22] and the editor of ontologies Web-Protégé [12]. However, the first plug-in is available only for participants of the Neon project, and the second can only be used in the WebProtégé editor, which is not very popular among developers of ontologies because of its limited functionality (compared to the desktop version of Protégé, for which, by the way, a similar plug-in is not developed).

The third problem is that patterns are described and applied separately and do not form a single system, making them difficult to share.

Another problem is that there are no widely available patterns that could be used in the development of ontologies of scientific subject areas.

2. METHODOLOGY OF DEVELOPMENT OF ONTOLOGIES OF SCIENTIFIC SUBJECT AREAS

The methodology discussed in this article [26] supports the construction of SSA ontology on the basis of the basic SSA ontology, containing descriptions of the most common entities characteristic of most scientific subject areas. Since the ontology of any SSA contains not only descriptions of its inherent system of concepts, tasks solved, and methods used but also descriptions of relevant information resources, the basic ontology of SSAs can be divided into four ontologies: (1) ontology of scientific knowledge, (2) ontology of scientific activity, (3) basic ontology of tasks and methods, and (4) basic ontology of information resources.

The ontology of scientific knowledge contains classes that set up structures for describing scientific concepts that are part of any SSA. Such concepts are *the branch of science*, *the object of research*, *the subject of research*, *the method of research*, *the scientific result*, *task*, *algorithm*, etc.

The ontology of scientific activity includes classes related to the organization of research activities, such as *person*, *organization*, *event*, *scientific activity*, *project*, and *publication*.

The basic ontology of information resources as the main class includes the class *information resource*, the set of attributes and relationships of which is based on the Dublin core standard [6]. On the basis of this class, relevant information resources modeled by SSAs can be described: databases, ontology, collections of documents, websites of organizations, persons, and projects.

The basic ontology of tasks and methods includes classes and relationships by which the tasks solved in this SSA, the methods of their solution, and the software components and algorithms that implement them can be described.

All basic ontologies have specifications in the OWL language [1]. For the most important concepts of the basic ontology of SSAs, content patterns have been developed, also implemented in the OWL language.

Since this methodology supports the development of ontologies by knowledge engineers and experts, it offers a system of ODPs that includes three types of patterns: presentation patterns, structural patterns, and content patterns.

Presentation patterns set the rules adopted in this methodology for naming and annotating the elements of ontology, similar to the rules adopted in the community of ontological modeling [16].

Structural logic patterns are designed to represent complex entities and structures relevant to the construction of ontologies of SSAs and for the description

of which there are no suitable expressive means in the OWL language. In particular, such patterns are designed to represent the domains of permissible values of arguments and *n*-ary and attributed relations (binary relations with attributes) [25].

The most important role in the methodology under consideration is played by content patterns, which are used for a uniform and consistent representation of the concepts used in SSAs and their properties. As mentioned above, such patterns are designed for concepts present in most SSAs. Owing to this, the development of ontology of a particular SSA is mainly reduced to the specialization (tuning) of content patterns in this area and building on their basis fragments of target ontology.

The construction of ontology of a specific SSA using basic ontologies and the system of ODPs is carried out in two stages:

1. Construction of components of ontology of the SSA on the basis of basic ontologies through their completion and development. At this stage, the specialization of content patterns and structural logical patterns presented in basic ontologies is performed on a specific SSA.

2. Population of the ontology of the SSA by concretization (instantiation) of content patterns and structural logical patterns presented in basic ontologies or derived from them by their specialization to a specific SSA.

At the same time, the ontology of the SSA is based on the patterns presented in the basic ontologies of scientific knowledge and scientific activity, the ontology of tasks and methods is based on the patterns of the basic ontology of tasks and methods, and the ontology of scientific Internet resources is based on the patterns of the basic ontology of information resources.

The specialization of an ontology design pattern is to rename it, add new properties (attributes and relationships) to it, and/or refine the property names and their value areas already described in the pattern.

Concretization of the ontology design pattern consists in substituting specific property values into it and adding the resulting fragment of ontology to the ontology being created.

In the development of ontology of a particular SSA, content patterns are mainly used, most of which are obtained by specialization of content patterns that form the basic ontology of SSAs.

As an editor of ontologies and ODPs at the moment, the popular editor Protégé is used. To support the use of ODPs, a data editor making it possible to populate the ontology of the SSA by specifying content patterns is provided.

3. EXPERIENCE OF USING THE METHODOLOGY IN THE DEVELOPMENT OF ONTOLOGIES OF VARIOUS SSAs

This methodology has shown its practical usefulness in the development of ontologies of various scientific subject areas: decision support [27], solving computationally complex problems of mathematical physics on supercomputers [10], plasma physics [20], etc. Consider the process of developing ontologies for some of them.

3.1. Development of Ontology of the Subject Area "Solving Computationally Complex Problems of Mathematical Physics on Supercomputers"

In the development of this ontology, the concepts of the basic ontology of scientific knowledge were clarified and detailed (*the object of research*, *the research method*, *the science section*, *the task*, *the algorithm*). At the same time, their specialization is fulfilled in this SSA. As a result, such entities as *the physical object* and *physical phenomenon*, *the fundamental law of nature*, *the physical model*, *the mathematical model*, *the system of equations*, *the numerical method*, *the parallel algorithm*, and others. Appropriate content patterns have been developed for these concepts.

Let us consider as an example the pattern for describing numerical methods (Fig. 1).

From a substantive point of view, such a pattern is the semantic neighborhood of the central concept, which in this case is the class of *numerical method*. For this concept, properties are defined, which are presented as attributes or relations. In this case, the attributes can be presented either as data properties—for those properties whose values have a standard data type (*Name, Description, Resistance to perturbations, Absolute accuracy, Order of accuracy, Computational complexity*)—or as object properties—for properties with values from the enumerated data type (*Solution type, Grid type, Discretization of the calculation area, Approach to the solution presentation*). In this case, the enumerated data types themselves are implemented in patterns to represent the areas of permissible values.

The relationships define the associations of objects of the class under consideration with objects of other classes and are presented as object properties. In particular, in this pattern, connections are set with the problems that this numerical method solves, with the parallel algorithm that implements it, with the systems of equations that it solves, etc.

Here is also a content pattern to describe parallel algorithms (Fig. 2).

This pattern defines the properties of a parallel algorithm. The central concept of this pattern is the *parallel algorithm* class. It also has properties whose values have a standard data type (*Name, Description, Scalability of evaluation*, etc.) and a property of *Decomposition of a calculation domain* with values from
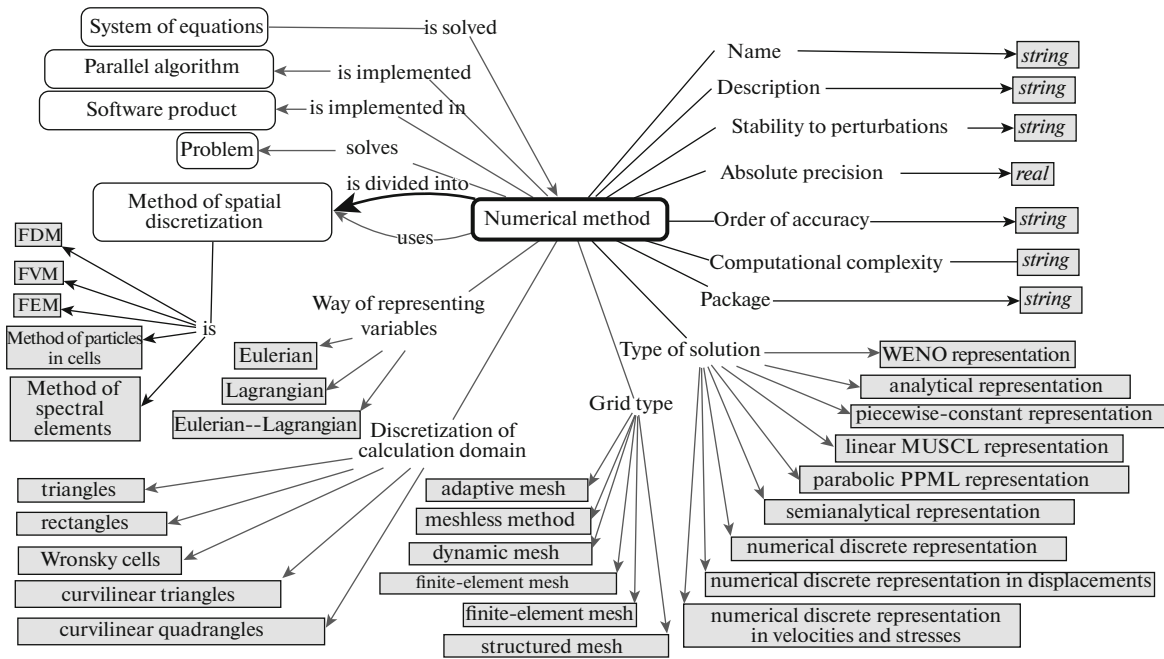
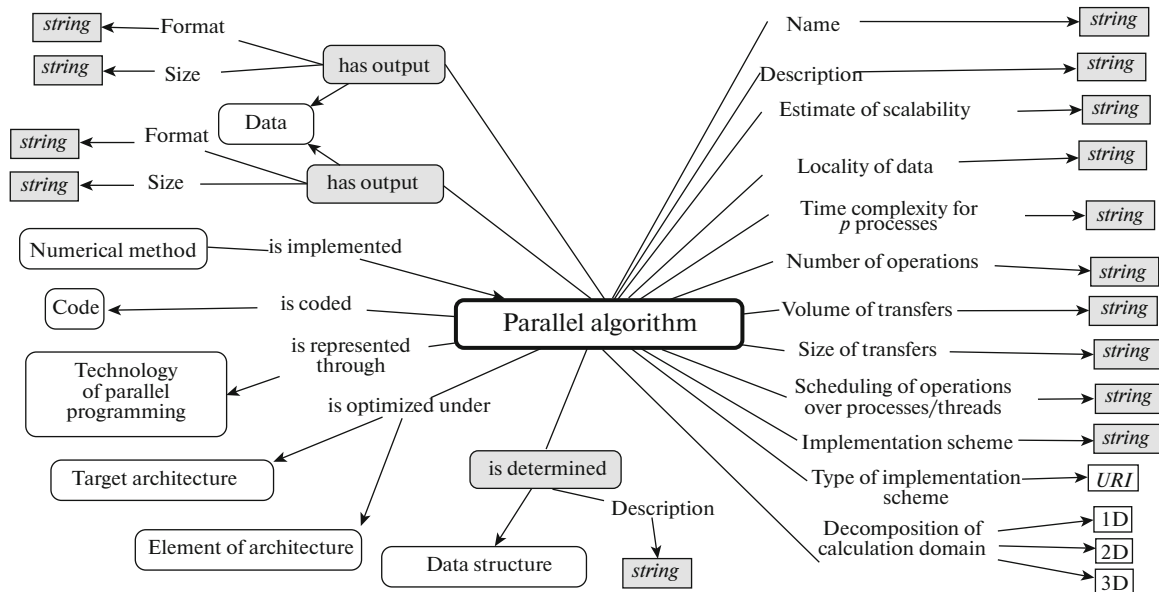**Fig. 1.** Content pattern of "numerical method."

Fig. 2. Content pattern for representing parallel algorithms.

the enumerated data type. In this pattern, a parallel algorithm can be associated with relationships to the target architecture, parallel programming technologies, the data structure used by it, its code, and the like. In this case, some relations can have their own attributes (*hasInput*, *hasOutput*, *isDetermined*).

Naturally, the patterns provided by the methodology may not be enough to describe this SSA in full. In this regard, it may be necessary to develop content patterns specific to a particular SSA. Such patterns for

this ontology are *target architecture*, *parallel programming technology*, *software product*, etc.

Figure 3 shows a pattern describing the concept of target architecture.

On the basis of the ontology discussed above, ontologies of several areas of mathematical physics were developed, such as astrophysics, geophysics, and plasma physics.

The content patterns presented in this ontology were used to describe these areas. Figure 4 presents the
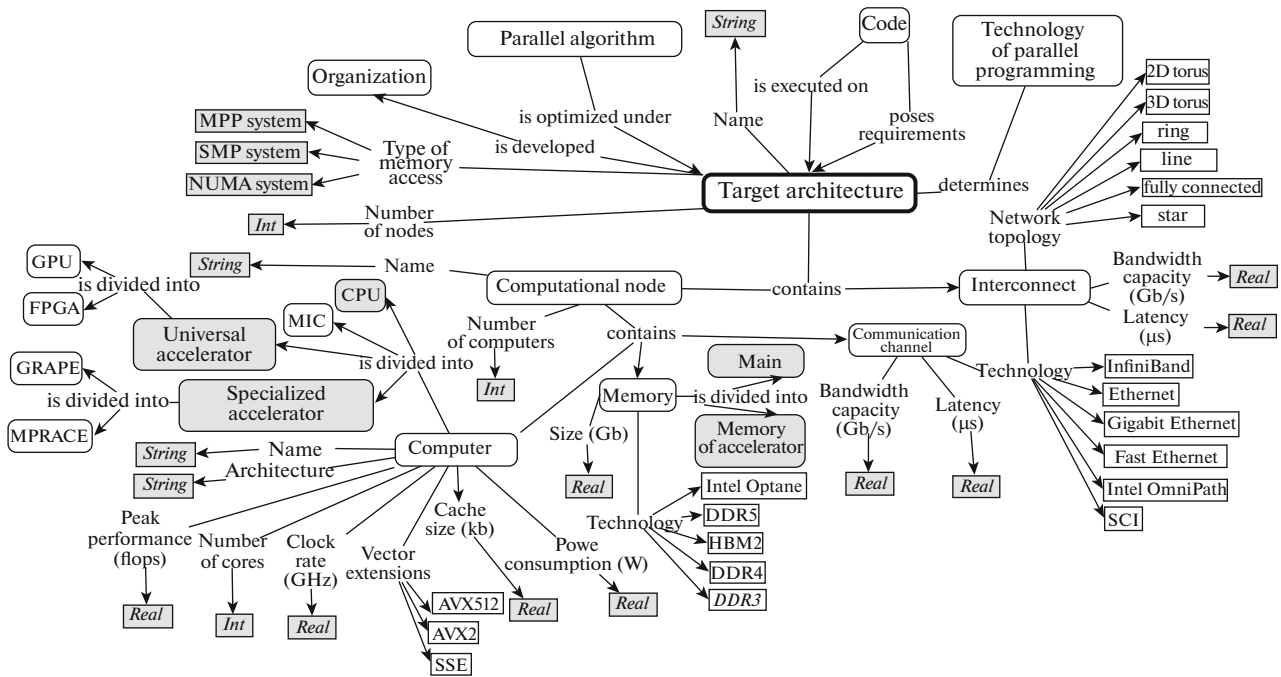
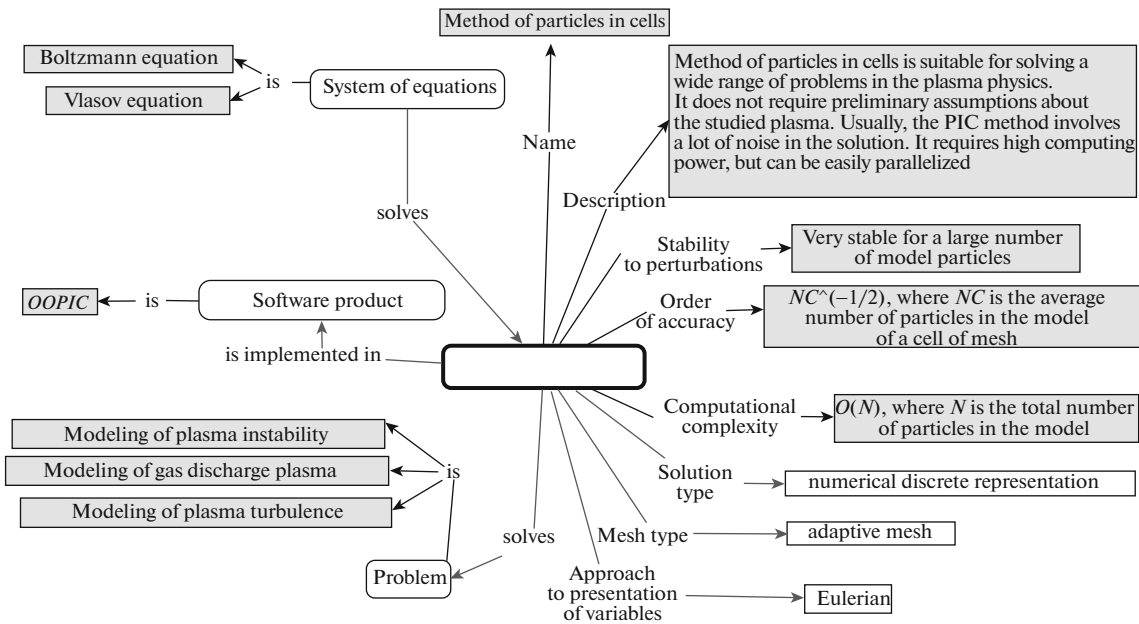**Fig. 3.** Content pattern of target architecture.

Fig. 4. Concretization of the content pattern of numerical method.

concrete definition of the pattern of numerical method to describe particle-in-cell method in the development of ontology of plasma physics [20].

3.2. Development of Decision Support Ontology

The peculiarity of this ontology is that it not only includes traditional concepts for scientific subject areas, such as methods, tasks, research objects, and scientific results, but also describes problem situations and stages of decision-making. It also reflects the orientation of information to different types of users and sets the connections of ontology objects with data from external sources that can be used in solving tasks of decision-making support (DMS). Along with the concepts related to the decision-making process and methods of its support, the subject area of DMS also considers issues related to the program implementation of methods, as well as activities carried out within its framework.
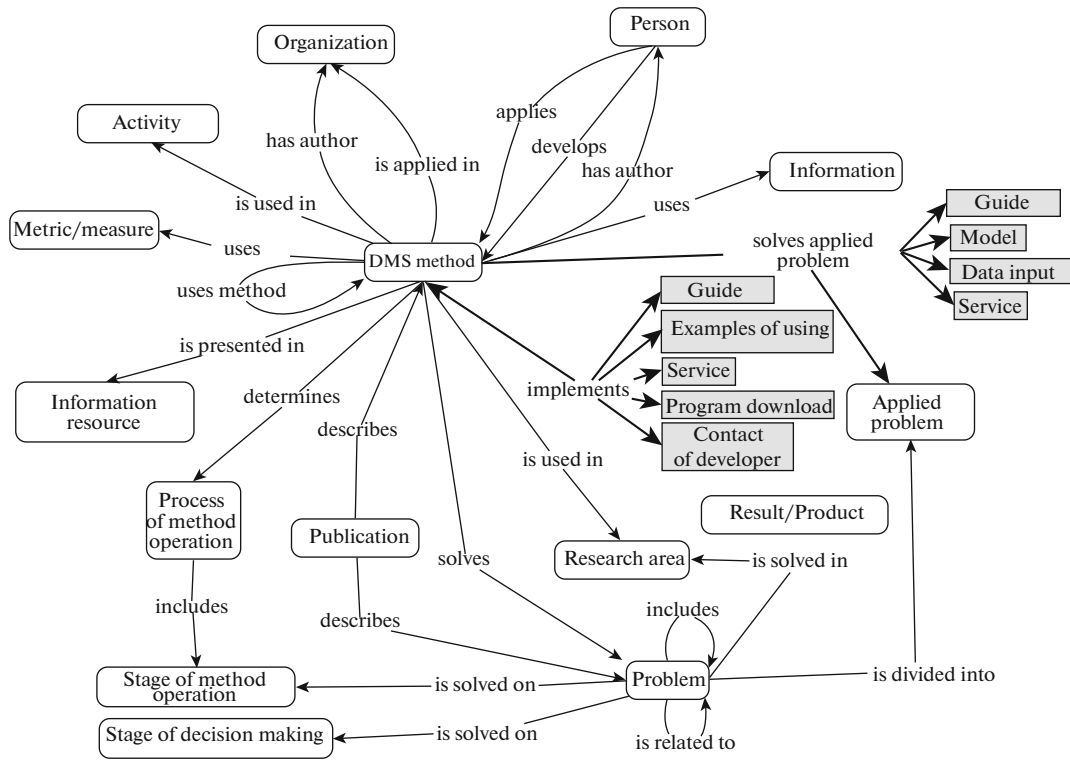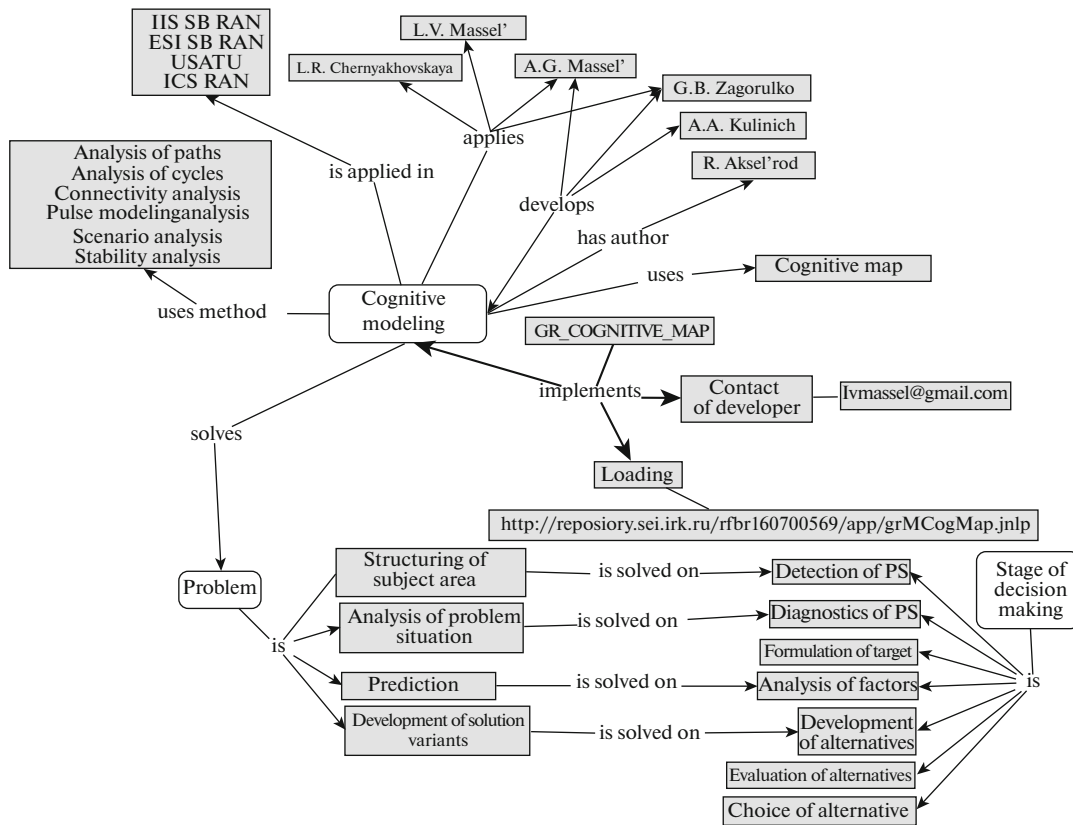
Fig. 5. Content pattern of DMS method.

Fig. 6. Cocnretization of the content pattern of "DMS method."

For concepts specific to this subject area (*problem situation*, *decision-making stage*, etc.), content patterns were developed from scratch. The main part of the content patterns used in the development of the ontology of the DMS was obtained by specializing of the content patterns, which are part of the basic ontology of the SSA.

Figure 5 shows a pattern for describing the DMS method obtained by specialization of the pattern "method of research" from the basic ontology of scientific knowledge.

The pattern of DMS method was used to add descriptions of DMS methods to the ontology. An example of specifying this pattern with information about the method of cognitive modeling is shown in Fig. 6.

From Fig. 6, it can be seen that the method of cognitive modeling uses cognitive maps, as well as such methods of path analysis, cycle analysis, and scenario analysis, and is designed to solve problems of structuring a subject area, analyzing a problem situation, developing solutions, etc.

CONCLUSIONS

Thus, our experience has shown the effectiveness of the use of ontological patterns of content in the development of ontologies of SSAs. This is due, in particular, to the fact that, in the ontology of any SSA, as a rule, a large number of typical fragments are present, which are well described by content patterns.

In addition, the use of content patterns allows us to provide a uniform and consistent representation of all the entities of SSA ontology, reduce the number of errors of ontological modeling, increase the "understanding" of ontology by developers, and thereby provide the possibility of collective development of ontologies.

Owing to the fact that the use of ontology design patterns greatly simplifies and facilitates the development of SSA ontology, it can be attractive to experts in the simulated field who do not have the skills of ontological modeling, which accelerates the development of SSA ontology.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. G. Antoniou and F. van Harmelen, "Web ontology language: OWL," in *Handbook on Ontologies*, Ed. by S. Staab and R. Studer, International Handbooks on Information Systems (Springer, Berlin, 2009), pp. 91–110. https://doi.org/10.1007/978-3-540-92673-3_4

2. M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, "A survey of ontology learning techniques and applications," Database **2018**, bay101 (2018). https://doi.org/10.1093/database/bay101

3. E. Blomqvist, K. Hammar, and V. Presutti, "Engineering ontologies with patterns: The eXtreme design methodology. Ontology engineering with ontology design patterns," in *Ontology Engineering with Ontology Design Patterns*, Ed. by P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, and A. Presutti, Studies on the Semantic Web, Vol. 25 (IOS Press, Amsterdam, 2016), pp. 23–50.

4. N. M. Borgest, "Ontology of designing a scientific direction: formation, development, examples," Ontologiya Proektirovaniya **12** (2), 136–157 (2022). https://doi.org/10.18287/2223-9537-2022-12-2-136-157

5. G. Brusa, M. L. Caliusco, and O. Chiotti, "Towards ontological engineering: A process for building a domain ontology from scratch in public administration," Expert Syst. **25**, 484–503 (2008). https://doi.org/10.1111/j.1468-0394.2008.00471.x

6. "DCMI Metadata Terms," http://dublincore.org/documents/dcmi-terms.

7. A. De Nicola and M. A. Missikoff, "A lightweight methodology for rapid ontology engineering," Commun. ACM **59** (3), 79–86 (2016). https://doi.org/10.1145/2818359

8. G. Ganino, D. Lembo, M. Mecella, and F. Scafoglieri, "Ontology population for open-source intelligence: A GATE-based solution," Software: Pract. Exper. **48**, 2302–2330 (2018). https://doi.org/10.1002/spe.2640

9. A. Gangemi and V. Presutti, "Ontology design patterns," in *Handbook on Ontologies*, Ed. by S. Staab and R. Studer, International Handbooks on Information Systems (Springer, Berlin, 2009), pp. 221–243. https://doi.org/10.1007/978-3-540-92673-3_10

10. B. M. Glinskiy, Y. A. Zagorulko, G. B. Zagorulko, I. M. Kulikov, A. F. Sapetina, P. A. Titov, and G. F. Zhernyak, "Building ontologies for solving compute-intensive problems," J. Phys.: Conf. Ser. **1715**, 012071 (2020). https://doi.org/10.1088/1742-6596/1715/1/012071

11. V. V. Gribova, S. V. Parshkova, and L. A. Fedorischev, "Ontologies for development and generation adaptive user interfaces of knowledge base editors," Ontologiya Proektirovaniya **12** (2), 200–217 (2022). https://doi.org/10.18287/2223-9537-2022-12-2-200-217

12. K. Hammar, "Ontology design patterns in Web-Protégé," CEUR Workshop Proc. **1486**, 50 (2015). https://ceur-ws.org/Vol-1486/paper_50.pdf.

13. O. P. Kuznetsov, V. S. Sukhoverov, and L. B. Shipilina, "Ontology as a systematization of scientific knowledge: structure, semantics, tasks," Tech. Software Tools Control, Monitoring Meas. Syst., 762–773 (2010).

14. O. P. Kuznetsov and V. S. Sukhoverov, "A ontological approach to determining the subject matter of scientific text," Ontologiya Proektirovaniya **6** (1), 55–66 (2016). https://doi.org/10.18287/2223-9537-2016-6-1-55-66

15. "NeOn Project," (2022). http://www.neon-project.org/nw/Welcome_to_the_NeOn_Project.

16. N. Noy and D. Mcguinness, *Ontology Development 101: A Guide to Creating Your First Ontology: Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880* (2001).

17. "Association for Ontology Design & Patterns," (2022). http://ontologydesignpatterns.org.

18. "Ontology Design Patterns (ODPs) Public Catalog," (2009). http://odps.sourceforge.net.

19. A. Sattar, E. Salwana, M. Nazir, M. Ahmad, and A. Kamil, "Comparative analysis of methodologies for domain ontology development: A systematic review," Int. J. Adv. Comput. Sci. Appl. **11** (5), 99−108 (2020). https://doi.org/10.14569/ijacsa.2020.0110515

20. A. V. Snytnikov, B. M. Glinskiy, G. B. Zagorulko, and Y. A. Zagorulko, "Ontological approach to formalization of knowledge in computational plasma physics," J. Phys.: Conf. Ser. **1640**, 012013 (2020). https://doi.org/10.1088/1742-6596/1640/1/012013

21. S. Singh and M. Aswal, "Ontology learning procedures based on web mining techniques," in *Int. Conf. on Advances in Engineering Science Management & Technology (ICAESMT) - 2019, Dehradun, India, 2019* (Elsevier, 2019). https://doi.org/10.2139/ssrn.3382660

22. M. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, "The NeOn Methodology framework: A scenario-based methodology for ontology development," Appl. Ontology **10**, 107−145 (2015). https://doi.org/10.3233/ao-150145

23. N. I. Tsukanova, *Ontological Model of Knowledge Representation and Organization: Textbook for High Schools* (Goryachaya Liniya-Telekom, Moscow, 2015).

24. Yu. A. Zagorulko, G. B. Zagorulko, and O. I. Borovikova, "Technology for building subject-based intelligent scientific internet resources based on ontology," Programmnaya Inzh. **7** (2), 51−60 (2016). https://doi.org/10.17587/prin.7.51-60

25. Yu. Zagorulko, O. Borovikova, and G. Zagorulko, "Development of ontologies of scientific subject domains using ontology design patterns," in *Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2017*, Communications in Computer and Information Science, Vol. 822 (Springer, Cham, 2018), pp. 141−156. https://doi.org/10.1007/978-3-319-96553-6_11

26. Yu. Zagorulko, G. Borovikova, and O. Zagorulko, "Pattern-based methodology for building the ontologies of scientific subject domains," in *New Trends in Intelligent Software Methodologies, Tools and Techniques, Granada, Spain, 2018*, Ed. by H. Fujita and E. Herrera-Viedma, Frontiers in Artificial Intelligence and Applications, Vol. 303 (IOS Press, 2018), pp. 529−542. https://doi.org/10.3233/978-1-61499-900-3-529

27. Yu. Zagorulko and G. Zagorulko, "Application of ontology design patterns for building an ontology of decision support in weakly formalized domains," CEUR Workshop Proc. **3044**, 13 (2021). https://ceur-ws.org/Vol-3044/paper13.pdf.

28. Yu. A. Zagorulko, "Automation of the development of ontologies of scientific subject domains based on ontology design patterns," Ontologiya Proektirovaniya **11** (4), 500−520 (2021). https://doi.org/10.18287/2223-9537-2021-11-4-500-520

Yury Alekseevich Zagorulko (born 1957) graduated from the Novocherkassk Polytechnic Institute in 1979, PhD (1989). He is the head of the Laboratory of Artificial Intelligence at the Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Science (Novosibirsk), Associate Professor at Novosibirsk State University. He is a Member of the Russian and European Associations for Artificial Intelligence. He is the author of more than 270 peer-reviewed publications in the fields of AI, knowledge and ontology engineering, intelligent system development, and computational linguistics. Author ID (Scopus): 23394231500; Researcher ID (WoS): R-1826-2016.

Galina Borisovna Zagorulko (born 1963) graduated from the Novosibirsk State University in 1985, PhD (2021). She is a senior researcher of the Laboratory of Artificial Intelligence at the Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Science (Novosibirsk), Associate Professor at Novosibirsk State University. She is a member of the Russian and European Associations for Artificial Intelligence. She is the author of more than 120 peer-reviewed publications in the field of AI, intelligent system development, knowledge engineering, ontological modeling and decision-making support. Author ID (Scopus): 35249924100; Researcher ID (WoS): K-2472-2018.