Building ontologies for solving compute-intensive problems

B M Glinskiy1, Y A Zagorulko2, G B Zagorulko2, I M Kulikov1, A F Sapetina1, P A Titov1 and G F Zhernyak1

1 Institute of Computational Mathematics and Mathematical Geophysics SB RAS, 6 Akad. Lavrentieva pr., Novosibirsk, Russia
2 A.P. Ershov Institute of Informatics System SB RAS, 6 Akad. Lavrentieva pr., Novosibirsk, Russia

E-mail: gbm@sscc.ru

Abstract. The aim of the study, results of which are presented in this paper, is to analyse methods and tools for constructing an ontology related to solving compute-intensive problems and to form algorithms of its use. This problem arises from the need to solve problems using modern and future supercomputers, containing millions and, in the long term, billions of simultaneously operating computing cores and having a huge degree of parallelism. In solving such problems, the researcher should be well versed in both computational methods for solving the problem and modern supercomputer technologies, which is not always the case. One of the solutions to this problem is the creation of a knowledge base that includes ontological descriptions of methods for solving compute-intensive problems and architectures of supercomputers that can be used to solve them. The development of ontologies for a given subject area is one of the most important stage in creating an intelligent support system for solving specific compute-intensive problems. The paper discusses the methods and tools that are used to build the ontology. The paper also presents examples of the development of ontologies for astrophysics and geophysics problems.

1. Introduction

The current level of development of computer technology and computational methods poses an important problem of the interaction of researchers with a supercomputer in solving compute-intensive problems. There is a need to use artificial intelligence methods for organizing such interaction and creating a kind of bridge between the formulation of the problem and methods for its solving, on the one hand, and applying the optimal supercomputer architecture, on the other. One of the possible approaches to solve this problem is described in [1, 2]. These papers consider the methods and tools that can be used to intelligent support for solving compute-intensive problems of mathematical physics on supercomputers. The conceptual scheme of such support consists of a knowledge base (KB), an information-analytical Internet resource (IAIR) on support for solving compute-intensive problems on supercomputers, an expert system, a library of program components, and a simulation module.

The core of the technology of support for solving mentioned above problems is the knowledge base, which includes a top-level ontology, an ontology of computational methods and parallel algorithms, an ontology of parallel architectures and technologies, as well as inference rules extending the logic of these ontologies.
Ontologies are currently the basis of KB of intelligent systems. They allow us to present the described areas in a structured way (as classes and their relationships) understandable to humans and available for computer processing.

In paper [3], the advantages of using ontologies in KB designing are described:

- Ontologies are a powerful cognitive tool.
- Ontologies make visible the semantic networks of human memory, as they present information in the form of a network of related concepts.
- Ontologies provide a systematic approach to the study of the subject area, which greatly facilitates its structuring.

It is also important to note the ontologies are a platform and subject-independent means for organizing interaction between various computer systems.

The ontology development is well supported by methodology and tools. There are some languages and software systems for ontology development [4–8]. The methods for construction [3, 9] and the evaluation of ontologies [10] are proposed.

It should be noted that information resources based on ontological descriptions are currently being created to support research in various scientific areas (physics [11], geology [12], biology [13], astrophysics [14, 15], etc. Web-resources on astrophysics [16], genomics [17], geology and geophysics [18] exist already.

The approach closest to the approach developed by the authors is described in [19]. It uses ontologies and inference mechanisms to support users solving compute-intensive problems on heterogeneous computing architectures, in particular on clusters equipped with NVIDIA GPGPUs and Xeon-Phi coprocessors complementing a traditional Intel processor. This approach allows us to find close to optimal solutions by combining various computing hardware, software, and planning strategies. This approach is demonstrated by solving the problems of bioinformatics and building fiber-optic networks.

The paper discusses the problems, methods, and tools for constructing ontologies intended for intelligent support for solving compute-intensive problems of mathematical physics on supercomputers. Examples of the development of ontologies for such fields as astrophysics and geophysics are given.

2. Problems and tools for building ontologies

The development of ontology is a complex creative task, the solution of which raises many problems. Many of these problems are typical, they are well known to knowledge engineers, and various methodological and software tools have been proposed to solve them. Other problems arise and are solved depending on the subject and problem areas of the designed ontology.

Consider the main problems of ontological engineering and the tools for constructing ontologies that allow us to solve them (figure 1).

At the initial stages of ontology construction, it is necessary to determine the boundaries of the modeled subject area (SA) and to highlight the most significant concepts and their properties. The problem area should be determined as well, i.e. the purpose of the created ontology should be taken into account since it is impossible to build a practically significant ontology in isolation from the context in which it will be used. This raises the problem of the completeness of the ontology. The ontology should describe SA in the volume that is necessary and sufficient for its practical applicability. To solve these problems, such methods of acquiring knowledge as methods for extracting knowledge from experts (questionnaires, interviews, lectures, brainstorming, round tables, etc.) and methods of working with the literature [3] are used. In addition to these methods, automated methods for collecting relevant information resources from the Internet and methods for extracting the SA concepts from them [20], as well as methods and tools for constructing subject dictionaries are also used.

When ontology developing, the use of base ontologies including the concepts of top-level, as well as ontology design patterns (ODPs), can be of great help. ODPs are documented descriptions of
proven in practice solutions for typical problems of ontological modeling. They are created for avoiding developers the common mistakes of ontological modeling [21]. The use of the base ontologies and patterns helps to solve the problem of ontology correctness by offering a scheme (framework) of ontology and consistent pieces of knowledge for its creation.

**Figure 1.** Ontological engineering problems and tools for their solution.

An effective software tool for consistency control is the reasoners [22], which operate based on axioms describing the properties of ontology classes. There are the reasoners embedded in some ontology editors and that allows us to control the correctness of the ontology in the process of its creation. The most famous and popular ontology editor is Protégé [7].

A good ontology should be well readable. To increase the perception of ontology by humans, various visualization methods and tools (specialized graphic editors, as well as plugins integrated into ontology editors) are used. They can represent ontology in the form of a graph with ontology classes as the nodes and relations between them as the arcs [7, 23]. Taking into account the cognitive characteristics of a human and observing the methodological recommendations in constructing a graphical representation of ontology (the number and balance of graph nodes, length of branches, etc.) contributes to its perception [3].

Another problem related to ontology design arises when third-party ontologies are used. In this case, we should integrate several ontologies. The solution of this problem requires a set of subtasks realization such as comparing, aligning, mapping, merging ontologies as well as extracting from them fragments necessary for building target ontology. The techniques for performing these manipulations with ontologies [24] and ontology editors providing tools to support some of them [7] are described in many publications.

Ontological modeling algorithms are described in [3, 25]. Their authors use different terminology for describing the process of constructing ontologies, focus on its various stages, and consider it with varying degrees of detail. Abstracting from specific techniques and tools, we can distinguish the following sequence of actions of a knowledge engineer in ontology building:

1. Decomposition of SA, allocation and identification of SA objects, formation of the SA glossary.
2. Classification of glossary concepts, formation of the meta-concepts (ontology construction from bottom to top), and specification of the concepts (ontology construction from top to bottom).
3. Description of the properties of concepts included in the glossary, the establishment of relationships between concepts, their visualization and formalization.

4. The population of the ontology with objects (instances of concepts), assignment of values of the objects properties, establishment of relationships between objects.

In the process of constructing ontology, there is often a need for its reengineering. At any step, it may be required a refinement of the elements of the ontology, resolution of contradictions, elimination of redundancy, restructuring and extension of the ontology.

3. Development of ontology for solving problems of mathematical physics on supercomputers

Consider the process of constructing the ontology of the subject area "Solving compute-intensive problems of mathematical physics on supercomputers", as well as the methods and tools that were used for this.

1. As a result of the interaction of specialists in the fields of mathematical physics, the use of supercomputers and ontological engineering, the main entities of the modeled domain were defined.

2. A multilevel classification of defined entities has been performed. First, top-level concepts such as Parallel Programming Technology, Target Architecture, Research Method/Tool, Software Product, etc. were described. Then these concepts were considered in more detail.

3. The properties of concepts and the relations between them are defined. At this stage, conceptual maps that represent the concepts under consideration, their properties and relations in the form of a graph were constructed using the tools of the Camp editor [23]. Based on these maps, a set of ontology design patterns describing fragments of ontology was developed. These patterns were then used in the development of an ontology for different subject areas (astrophysics, geophysics, plasma physics, etc.). Figure 2 shows a pattern that describes the “Target Architecture” concept.

**Figure 2.** A pattern for describing Target Architecture.
Computing node and Interconnect contained in Target Architecture have their data properties (attributes) and object properties (relations) (figure 2). Some of these attributes have the values of standard data types. And such attributes as Vector extensions, Technology, Network topology have the values of enumerated types (domains) represented in the ontology by the corresponding classes. For the description of domains, special structural patterns are designed [21].

4. After constructing the class system, identifying and formalizing their properties (attributes and relations), the instances of the classes (the specific objects in the given area) were created, the values of their attributes were defined, and the relations between them were established. When creating a specific object, we used a pattern that describes its class which is like the pattern shown in figure 2.

Since the developed ontology should become the basis of the intelligent support system, it should be presented in a form adapted for computer processing. To formalize the ontology, the OWL language [6] and the Protégé editor [7] were used. In figure 3 the Class hierarchy field shows the basic concepts of the ontology being developed, the Usage field shows the properties of the Target Architecture class, and the Description field presents an axiom stating that the Target Architecture contains Computing nodes and Interconnect.

Figure 3. Properties of class Target Architecture.

Ontology objects connected to each other form a semantic network which can be used for navigating from one object to another. For example, in the context of support for solving compute-intensive problems on supercomputers, it is possible to build a chain from the problem investigated by the user to parallel architectures, technologies, and codes that can be used for the solution of this
problem. In addition, using the semantic network, the inference engine can make a logical conclusion and deduce facts that are not explicitly presented in the ontology.

Let us consider such semantic chains by an example of the problems of astrophysics and geophysics.

4. Examples of astrophysical and geophysical ontologies and navigation through it

The ontology fragments presented in this section were developed by the authors based on the top-level concepts and ontology design patterns described above.

The first ontology has developed for solving astrophysics problems [1]. A fragment of this ontology and a semantic chain for solving the problem of modeling the development of hydrodynamic turbulence of the interstellar medium at the early stage of evolution in the maximum resolution for the qualitative determination of gas fragmentation are presented in figure 4. The links of this chain are highlighted in bold and in color.

![Ontology Diagram](image)

**Figure 4.** The ontology for solving compute-intensive astrophysics problems with a semantic chain for solving a real problem.

We fix the Target Architecture on which the calculations are planned. Let it be an MPP-system equipped with Intel Xeon Phi processors of the Knights Landing generation (KNL). Such systems are installed, for example, on the base of the Siberian Supercomputer Center of ICMMG SB RAS. To describe the process of interstellar turbulence development at an early stage when the nature of the
velocity perturbation is more significant than the pressure – gravity interaction, the equations of magnetic isothermal hydrodynamics without gravity is suitable. For the correct numerical solution of the MHD equations, we can choose the Godunov method implementation based on the fifth-order WENO scheme. The ability of KNL to carry out arithmetic operations simultaneously on a vector of length 512 bits (AVX512) allows us to efficiently implement the selected Numerical Method. For parallelization between cluster nodes, we take the most optimal 3D decomposition of the computational domain. The fixed Target Architecture specifies the Parallel Programming Technologies that need to be used: MPI to exchange between nodes, OpenMP to enable multi-core, and the AVX512 extension for code vectorization. That way, a scheme has defined that allows us to obtain a scalable parallel code that numerically solves the problem on the selected computational architecture with a high order of accuracy.

Another example relates to solving geophysical problems. Figure 5 shows a fragment of the ontology and a semantic chain for solving problems of modeling 3D wave field propagation in isotropic media with curvilinear surface topography.

Figure 5. The ontology for solving compute-intensive geophysical problems with a semantic chain for solving a real problem.
In solving the problem, the finite difference method in space and time is applied. The mathematical model of the simulated process is represented by a dynamic linear elasticity system, while the medium is assumed to be isotropic. One of the most developed solution methods is the finite difference method. The solution in this case is presented in numerical discrete form. The system itself is represented in displacements. Domain discretization (partition into units) is selected in the form of curved quadrangles for consistency with curved surface topography. Due to the large volumes of required computing power and RAM, a parallel algorithm and 3D decomposition of the computational domain are used to implement the explicit scheme of the selected method.

The parallel algorithm is optimized for calculations on computing devices (computers). For example, in the scheme represented in figure 5, computers of the Intel Xeon Phi family are used (KNL type). The specificity of this device and MPP memory access type imply the use of parallel programming technologies: MPI - for organizing data exchanges between computers and AVX - for code vectorization inside each computer. This way one can guarantee effective use of the finite difference method when dealing with large data.

5. Conclusions
The paper considers the problems of ontological engineering, methods, and tools for ontology construction. It is considered how these methods and tools were used for intelligent support of solving compute-intensive problems on supercomputers. The system of top-level concepts and the patterns of ontology design was developed. They were used in the construction of ontologies for solving problems of astrophysics and geophysics, as well as for defining specific objects from these areas.

A set of related objects forms a semantic network that allows us to navigate through the ontology and to reveal semantic chains between objects that do not have explicit connections. The paper gives examples of constructing semantic chains linking the tasks of astrophysics and geophysics with parallel algorithms, technologies, and architectures.

Acknowledgements
This research was supported by the Russian Foundation for Basic Research [grants No. 19-07-00085 and No. 19-07-00762].

References
[1] Glinskiy B, Zagorulko Y, Zagorulko G, Kulikov I and Sapetina A 2019 The Creation of Intelligent Support Methods for Solving Mathematical Physics Problems on Supercomputers Communications in Computer and Information Science 1129 427–38
[2] Zagorulko G, Zagorulko Y, Glinskiy B and Sapetina A 2019 Ontological Approach to Providing Intelligent Support for Solving Compute-Intensive Problems on Supercomputers Communications in Computer and Information Science 1093 363–75
[3] Gavrilova T and Muromtsev D 2008 Intelligent technologies in management: tools and systems (St. Petersburg: Publ. house of St. Petersburg State University) p 488
[4] Uschold M 1996 Building Ontologies: Towards a Unified Methodology Technical Report (Edinburgh, UK: University of Edinburgh)
[5] Corcho et al 2003 Data & Knowledge Engineering 46 41–64
[6] Antoniou G and Harmelen F 2004 Web Ontology Language: OWL Handbook on Ontologies ed S Staab and R Studer (Berlin: Springer Verlag) pp 67–92
[7] Protégé A free, open-source ontology editor and framework for building intelligent systems Online: http://protege.stanford.edu (last visited: 20.05.2020)
[8] Sure Y, Staab S and Studer R 2009 Ontology Engineering Methodology Handbook on Ontologies ed S Staab and R Studer (Berlin: Springer Verlag) pp 135–52
[9] Zagorulko Y, Borovikova O and Zagorulko G 2017 Methodology for the development of ontologies for thematic intelligent scientific Internet resources Pro. 2nd Russian-Pacific Conf. on Computer Technology and Applications (RPC-2017) (Vladivostok) pp 194–8
[10] Garilova T, Bolotnikova E and Gorovoy V 2012 New Ergonomic Metrics for Educational Ontology Design and Evaluation *New Trends in Software Methodologies, Tools, and Techniques Frontiers in Artificial Intelligence and Applications* series Proc. 10th SoMet_12 ed H Fujita and R Revetria (IOS Press) vol 246 pp 361–78

[11] Cvjetkovic V 2017 Web physics ontology: Online interactive symbolic computation in physics *Proc. 4th Experiment@International Conf. (exp.at’17)* pp 52–7

[12] Xiaogang M 2011 *Ontology Spectrum For Geological Data Interoperability* (ITC Netherlands) p 204

[13] Cook D, Neal M, Bookstein F and Genna R 2013 Ontology of physics for biology: representing physical dependencies as a basis for biological processes *J Biomed Semantics* 4 41

[14] Sarro L and Martínez R 2003 First Steps towards an Ontology for Astrophysics *KES, LNCS* ed V Palade et al (Heidelberg: Springer) vol 2774 pp 1389–95

[15] Louge T, Karray H, Archimède B and Knödlseder J 2015 Semantic Interoperability in Astrophysics for Workflows Extraction from Heterogeneous Services *Lecture Notes in Business Information Proc.* p 213

[16] ESPAS. Near earth space data infrastructure for e-science Online: https://www.espas-fp7.eu/portal/browse.html#ontology (last visited: 20.05.2020)

[17] SemGen. Experimental software tool for automating the modular composition and decomposition of biosimulation models Online: http://sbp.bhi.washington.edu/projects/semgen (last visited: 20.05.2020)

[18] Awesome geoscience semantics. Online: http://www.geoscience-semantics.org/refs/ (last visited: 24.05.2020)

[19] Faheem H, König-Ries B, Aslam M, Aljohani N and Katib I 2018 Ontology Design for Solving Computationally-Intensive Problems on Heterogeneous Architectures *Sustainability* 10(2) 441

[20] Akhmadeeva I, Zagorulko Y and Mouromtsev D 2016 Ontology-Based Information Extraction for Populating the Intelligent Scientific Internet Resources *J Communications in Computer and Information Science* 649 119–28

[21] Zagorulko Y, Borovikova O and Zagorulko G 2018 Pattern-Based Methodology for Building the Ontologies of Scientific Subject Domains *New Trends in Intelligent Software Methodologies, Tools and Techniques. Proc. 17th Int. Conf. SoMeT_18* ed H Fujita and E Herrera-Viedma (Amsterdam: IOS Press) Series: Frontiers in Artificial Intelligence and Applications vol 303 pp 529–42

[22] Abburu S 2012 A Survey on Ontology Reasoners and Comparison *J Computer Applications* 57(17) 33–9

[23] IHMC CmapTools Free toolkit for building conceptual knowledge maps Online: http://cmap.ihmc.us/ (last visited: 02.06.2020)

[24] Noy N and Musen M 1999 SMART: Automated Support for Ontology Merging and Alignment *Proc. 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW’99)* (Banff Canada)

[25] Gorshkov S 2016 Introduction to ontological modeling TriniData LLC Online: http://trinidata.ru (last visited: 20.05.2020)