ST. PETERSBURG NATIONAL RESEARCH RESEARCH UNIVERSITY OF INFORMATION TECHNOLOGY, MECHANICS AND OPTICS

As a manuscript

Razdiakonov Daniil Sergeevich

METHOD FOR VISUAL MANAGEMENT OF ONTOLOGICAL DATA IN THREE-DIMENSIONAL VIRTUAL SPACE

SUMMARY

of the dissertations for an academic degree
PhD in Computer Science

Academic supervisor:
Candidate of Technical Sciences, Associate Professor
Muromcev Dmitriy Iliych

Saint Petersburg — 2024

Summary

Relevance. Currently, more and more software systems include components designed to work with semi-structured data. To store and process data, such systems use graph formats and algorithms, and ontologies are used as schemes or data models. The organization of human-machine interaction in such systems by means of standard forms and tables is not efficient enough, since poorly structured information does not imply a fixed user interface for display, therefore, various visualization formats, including interactive ones, are widely used in such systems.

According to the work «About information visualization» [0], the human visual system is able to effectively perceive data presented in a graphical form, that is, visualization helps to identify images, build hypotheses and extract patterns from data, which makes it more efficient. the so-called process of data mining, or, in other words, iterative building of queries to systems and analysis of the results obtained. It is important to note that the data mining task is highly dependent on the type of data and the development of interfaces for its solution cannot be performed using a single template.

Ontology-based systems are used in many areas, such as bioengineering, energy industry, instrumentation, biology, utility accounting systems, cultural heritage, etc. Currently, there are several approaches to visualizing ontologies. These can be UML or node-link diagrams, or more specialized systems that allow you to form an ontological graph, for example, the WebVOWL [0] or Ontodia [0] tools. Many industrial data visualization tasks are solved using virtual reality systems. For the most part, objects of the real world act as data for such visualizations, direct interaction with which, for one reason or another, is impractical or impossible. Examples include training simulators for challenging environments that are time and resource intensive. The reproduction of such conditions by means of virtual reality glasses in the framework of the curriculum is much less resource-intensive, but at the same time sufficient to acquire the necessary skills. Another example would be the visualization of real-time data from a drone or remote spatial scanning system located at a significant distance from the operator. Another example is interaction with the microcosm through the visualization of particles and molecules, which is impossible in physical reality.

Existing virtual reality systems use pre-prepared fixed models and do not offer opportunities for working with semi-structured data, such as ontological knowledge bases. This fact is a significant limitation of virtual reality systems, since it does not allow dynamically connecting new data sources to the rendered models.

Various tools for working with data in virtual space are discussed in the articles «Immersive and Collaborative Data Visualization Using Virtual Reality Platforms» [0] and «Virtual Reality: Beyond Visualization » [0]. The first article discusses the possibilities of collaborating on data in virtual reality, as well as ways to express the properties of data, and in the second, the ways of using virtual reality in the scientific environment. However, the tools listed in the articles do not touch upon the topic of ontology visualization.

Examples of visualizations that partially use ontological data, for example, are applications developed by the British Museum, which conducts virtual tours, or applications of the American Museum Association, which provided the opportunity to visit museums in a pandemic using virtual reality technologies. At first glance, these examples are in no way related to the visualization of ontological data, but given the fact that museum data are stored in an ontological form, these types of visualization can be called a customized way of visualizing ontological data. Thus, the task of visualizing ontological data in virtual space is relevant along with other types of visual representations, but at the moment there are no ready-made solutions that allow working with ontological data in virtual reality. At the same time, such visualizations cannot be obtained by simple transformation of two-dimensional diagrams into three-dimensional ones due to the developed and rich semantics of visualization elements, which requires special methods and tools to form visual representations.

The problems in this area are: the complexity of the formation of queries to ontological knowledge bases through the management of three-dimensional visualization, which generates an excessive number of search parameters, the processing of blank ontology nodes, as well as the lack of objective mechanisms for evaluating the constructed visualizations. This work is devoted to the study of the listed problems.

State of the art. The topic of visualization and visualization of ontologies, in particular, is the subject of many works by scientists from all over the world. Visualization, for example, is devoted to the work of such Russian scientists as: M.M.

Matyushin, T.G. Vakurina, V.V. Kotel, P.A. Lomov, M.G. Shishaev, N.S. Vagarina, D.S. Marinin, Z. V. Apanovich, P.S. Vinokurov, T.A. Kislitsina, E.E. Kotova, I.A. Pisarev. At the same time, many foreign articles are being written on the topic of ontology visualization. Among the authors of these articles there are such scientists as: A. Katifori, C. Halatsis, G. Lepouras, C. Vassilakis, J. Sevilla, P. Casanova-Salas, S. Casas-Yrurzum, T.B.A. Vasyliuk, N.O.P. Vago, M. Sacaj, M. Sadeghi, S. Kalwar, M.G. Rossi, L. Asprino, C. Colonna, M. Mongiovì, M. Porena. The topic of visualization metrics was also touched upon in the works of the following Russian authors: M.V. Kolomeets, A.A. Pronoza, A.A. Chechulin, but is more common in the works of foreign authors. Foreign authors touching upon the topic of visualization metrics are such scientists as: C. Lewerentz, F. Simon, G. Melançon, A. Sallaberry, E. Bertini, A. Tatu, D. Keim, R. Brath, E. Bertini, G. Santucci.

The research aim of the work is to accelerate of interactive search in ontological knowledge bases. The following **tasks** were solved to achieve this goal:

1. Analytical review of methods for representing ontological data in three-dimensional space.
2. Development of a taxonomy of interactive visualization methods and metrics of search labor intensity to assess and substantiate the relevance of the developed method.
3. Investigation of the possibilities of using language models for vectorizing queries to knowledge bases and ranking visualization elements.
4. Investigation of the problem of incomplete coverage of data space in areas containing blank nodes.
5. Development of an experiment plan and a software prototype that implements the methods proposed in the dissertation and assembly of an experimental stand.
6. Carrying out tests confirming the relevance and effectiveness of the developed method.
7. Processing and generalization of experimental research results.

The object of research is human-machine interaction, interactive search in ontological knowledge bases and methods of visualization of ontological data.

The subject of research is visualization models, methods of constructing and evaluating interactive visualizations of ontological data, as well as algorithms for processing blank ontological nodes.

Research Methods used in the work: methods of statistical processing of experimental data, prototype design, graph theory, deep learning method, brainstorming, data vectorization, set theory, ontological engineering, testing and work with groups of subjects.

Defended research statements:

1. A method for visual management of ontological data based on an agent-based model of the interactive visualization process, which allows reducing the complexity of executing user scripts by reducing the step-by-step construction of visualizations to the task of searching in the state space.
2. Search effort metrics for assessing the quality of interactive visualizations, providing optimization of the process of their creation, independent of the type of tool used.
3. An algorithm for ranking interactive visualization elements that uses query vectorization and an attention mechanism to automatically calculate the context of visualizations based on the semantic proximity of its elements.
4. An algorithm for generating context-sensitive identifiers for anonymous nodes of an ontological graph, allowing to expand the data space available for visualization.

In this work, we combine the concepts of visualization and visualization control into the term «Interactive visualization», which, in response to user actions, changes the scale, viewing angle of the data, changing filters and highlighting the most significant fragments for the user, etc. For interactive visualizations, only partially the metrics generated for static visualizations are applicable. At the same time, the quality of interactive visualization, like the quality of visualization, is a subjective concept. In order to give an objective assessment of the quality of three-dimensional interactive representations of ontological data, a model was developed based on the identified use cases, and a smart highlighting mechanism was developed to improve certain aspects of interactive visualizations.

The scientific novelty of the research results is due to the following:

1. A method for developing tools for interactive visualization of ontological data is proposed based on a model that describes the target function of the interactive visualization process in terms of speed, labor intensity and distance, which allows reducing the labor intensity of constructing interactive visualizations of ontological data.

2. Metrics of the complexity of interactive visualization of a multidimensional data space have been developed, which differ in invariance to the user's control action on the visualization of the ontological graph.
3. A new application of the language model on the Transformer architecture for ranking vectorized nodes of an ontological graph is proposed, which makes it possible to reduce the complexity of searching in the data space.
4. A new application of the language model on the Transformer architecture is proposed for vectorization of queries automatically generated for given nodes of the ontological graph, which allows reducing the complexity of searching in the data space by sequentially clarifying the semantic context.

Author's personal contribution. The content of the thesis and the main provisions submitted for defense reflect the personal contribution of the author to the published work. The author personally developed an algorithm for lazy visualization of blank nodes of ontological graphs using context-sensitive identifiers, was directly involved in the study and implementation of vector representations of graph nodes in the Ontodia visualization tool. In the course of work on the dissertation, the author developed and investigated an algorithm for ranking the nodes of the visualized ontological graph, using vectorization of the request and the original graph, as well as the attention mechanism for ontological graph data. The author independently wrote the main substantive parts of the articles that formed the basis of this dissertation research, and also personally presented scientific reports on the topic of the dissertation at conferences. The author has developed and integrated the software written in the course of the research, namely the L3Graph software module and the Ontodia3d software product.

Implementation of the results. The results of this dissertation research were implemented in the research projects of the company «Metaphacts GMBH» and its St. Petersburg branch LLC «Metaphacts East Europe», including the development of an experimental extension of the platform «Metphactory» [0] to visualize ontological data in virtual space.

Credibility and validity of the results obtained within the framework of the dissertation work is confirmed by the apparatus for evaluating the research results formed at the initial stage of the work – the metrics of the labor intensity of the search. In particular, the reliability is confirmed by theoretical calculations proving the consistency of scientific results, as well as by the consistency of the experimental data obtained on the test bench developed by the author. The validity of the research results is confirmed by the approbation of the research results at Russian and international conferences, as well as their consistency with the previously obtained research results of other authors in this subject area.

Approbation of results. The results of the work were tested by international and Russian conferences. The list of conferences at which the results were presented is as follows:

1. ESWC 2019 – Extended Semantic Web Conference (2019).
2. KESW 2017 – Knowledge Engineering and Semantic Web (2017).
3. ESWC 2016 – Extended Semantic Web Conference (2016).
4. ISWC 2015 – International Semantic Web Conference (2015).
5. X КМУ ИТМО – Конгресс молодых ученых (Congress of Young Scientists).

The theoretical significance of the results obtained is to substantiate the expediency of taking into account the user's control action on visualization, as well as his focus of attention when constructing a model of interactive visualization of ontological data. In addition, methods of reducing the complexity of searching in the space of ontological data by ranking the vectorized nodes and the language model on the «transformer» architecture are shown.

The practical significance of the results of the dissertation work lies in the creation of the development of an experimental stand for visualization of ontological data in virtual space and the calculation of labor intensity metrics for interactive search by graph, as well as in the implementation of a software library for visualization methods using an algorithm for issuing context-sensitive identifiers and an attention mechanism for ordering nodes. A set of recommendations is proposed to improve interactive visualization in accordance with the metrics of the complexity of the search at different steps. In addition, the author trained a neural network for ordering nodes, which is used in the algorithm for issuing context-sensitive identifiers.

The value of scientific works is determined by the fact that for the first time a software architecture was proposed and a system for interactive visualization of ontological data in three-dimensional space was built. The proposed solutions open up new possibilities for the practical application of knowledge bases in virtual reality systems, both for educational or research tasks, and in production processes, where such technologies are increasingly used. Also, new algorithms are considered that allow to speed up interactive search in ontological knowledge bases due to context-sensitive processing of blank nodes and ranking of ontological graph elements based on query vectorization and attention mechanism. Metrics of the complexity of interactive visualization of a multidimensional data space invariant with respect to the user's control action are proposed. The proposed models and algorithms presented in scientific papers are of high practical and scientific value, which is confirmed by numerous presentations at conferences.

Articles on the Thesis Topic. The main results on the topic of the dissertation are presented in 8 printed publications, of which 6 publications in journals peer-reviewed by Web of Science or Scopus, 1 publications in journals from the list of the Higher Attestation Commission. Approbation of the research results has been confirmed by 3 public reports at all-Russian and/or international conferences over the past 3 years: KESW 2017, ESWC 2019, X CMU ITMO. Received 1 certificate of a computer program.

Scientific publications included in international abstract databases and citation systems:

1. Razd'Yakonov D., Morozov A., Pavlov D., Muromtsev D. Approach to Blank Node Processing in Incremental Data Visualization by the Example of Ontodia // Programming and Computer Software - 2020, Vol. 46, No. 6, pp. 384-396
2. Razdyakonov D., Wohlgenannt G., Emelyanov Y., Pavlov D., Mouromtsev D. A New Tool for Linked Data Visualization and Exploration in 3D/VR Space // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2019, Vol. 11762, pp. 167-171
3. Wohlgenannt G., Klimov N., Mouromtsev D.I., Razdyakonov D., Pavlov D., Emelyanov Y. Using word embeddings for visual data exploration with ontodia and wikidata // CEUR Workshop Proceedings - 2017, Vol. 1932
4. Wohlgenannt G., Klimov N., Mouromtsev D., Razdyakonov D., Pavlov D., Emelyanov Y. Using word embeddings for search in linked data with Ontodia // CEUR Workshop Proceedings - 2017, Vol. 1963, pp. 16-24
5. Mouromtsev D., Pavlov D., Emelyanov Y., Morozov A., Razdyakonov D., Parkhimovich O. Workflow supporting toolset for diagram-based collaborative ontology development implemented in the open budget

domain // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2016, Vol. 9989, pp. 178-182

6. Mouromtsev D., Pavlov D., Emelyanov Y., Morozov A., Razdyakonov D., Galkin M. The simple, web-based tool for visualization and sharing of semantic data and ontologies // CEUR Workshop Proceedings - 2015, Vol. 1486, pp. 77

Scientific publications included in the list of Russian peer-reviewed journals:

7. Раздьяконов Д.С., Морозов А.В., Павлов Д.С., Муромцев Д.И. Подход к обработке пустых узлов при порционной визуализации данных на примере инструмента ONTODIA // Программирование - 2020. - No 6. - С. 16-29

Introduction is devoted to the description of the main components of the dissertation work. Here the relevance of the work, the goal and the tasks that must be completed to achieve it are determined. The provisions for the defense are presented, and the scientific novelty of the proposed methods and algorithms is described. The first chapter provides a thorough overview of the subject area and provides definitions and descriptions of key terms, methods and algorithms. The dissertation work touches upon the areas of knowledge related to the visualization of ontological graphs and the metrics for evaluating the constructed visualizations. At the stage of constructing representations, language models are also used to improve perception and optimize metric values. The work also solves the problem of visualizing anonymous nodes, so this area of knowledge is also subject to analysis in the first chapter. Given the areas of research listed above, the chapter is structured accordingly. At the beginning of the chapter there is an overview of the current visualization metrics, including an overview of the metrics presented in the articles. «Empirical evaluation of aesthetics based graph layout» [0], «Shape-Based Quality Metrics for Large Graph Visualization» [0], «Quality metrics for information visualization» [0] и «Quality metrics in high-dimensional data visualization: An overview and systematization» [0].

The last section also provides examples of metrics for visualizing ontologies, which are described in the article «Developing 12 Non-Empirical Metrics and Tools for Ontology Visualizations Evaluation and Comparing» [0]. The next section of the first chapter is devoted to graph visualization tools. When performing a review of visualization tools, it was revealed that the number of visualization tools for ontological data in three-dimensional space is extremely small. Therefore, since the visualization tools for three-dimensional graphs can be adapted for visualizing ontological graphs, it was decided to include the visualization tools for three-dimensional graphs in this review. The visualization tools were reviewed according to the following criteria: the used layout algorithm, the amount of visualized data, navigation tools, data management tools. The following renderers were reviewed: «Plotly: 3D Network Diagram», «Cy3d», «Force Atlas 3D», «Mayavi-2010», «Walrus», «3D Networking Tools Visualization for Mac OS X», «KiNG Display Software», «3d-force-graph (from Vasco Asturiano)», «Graph-Visualization (from David Piegs)», «Graphosaurus (from Corey Farevel)», «Ngraph.pixel (from Andrey Kashka)», «DiVE (from NLeSC)», «OntoSphere3d». At the end of the section, there is a summary table of the visualization tools. The next section of the chapter is devoted to the problem of handling empty nodes (anonymous nodes). The material was discussed in detail in the articles «Approach to Blank Node Processing in Incremental Data Visualization by the Example of Ontodia» written as part of the dissertation work. The section ends with the definition of the mechanism of attention and its application in modern language models. It also provides a brief overview of various language models based on the Transformer architecture. An overview of the following language models is presented: PT-2, GPT-3, RuGPT-3, GPT-GNN, Microsoft T-NLG, Google mT5-XXL, Nvidia MegatronLM.

The second chapter of the thesis is devoted to the description of the model of interactive visualization of ontological data and the description of task-oriented metrics of the search complexity based on the developed model. The first part of the chapter defines the visualization taxonomy. The taxonomy is based on the ontological graph G, where G is a set of triplets of the format $<$s, p, o$>$. Graph G defines a multidimensional data space, where each dimension corresponds to a predicate from the ontology. Any data representation, according to our model, is a tuple of two elements: **data visualization function** and **control function**.

The final image that we see on the screen is usually not the entire data space of D, but the projection of part of the data space onto the flat screen of the monitor V_D. The area of data space that is displayed on the screen is called the SF search box, which has a position and width for each dimension of the D_i data space. Figure 1 shows a two-dimensional space with a search box in it, and figure 2

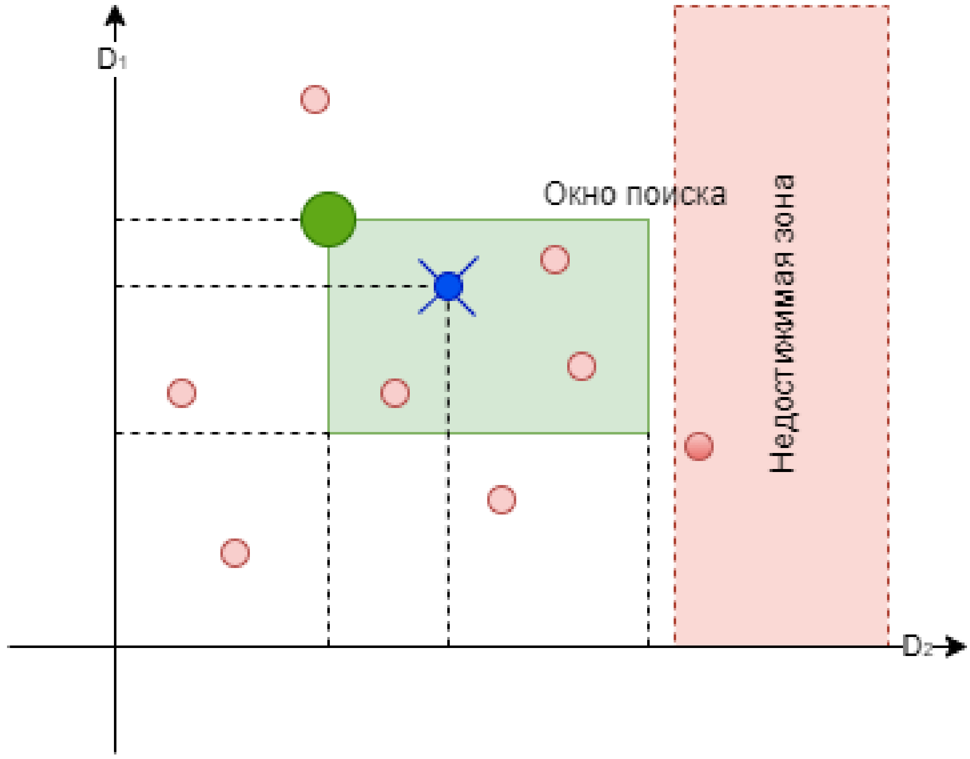

Рисунок 1 — 2D Data space

shows a three-dimensional space with a search box in it. The visualization function $V_D(SF)$ is responsible for calculating this projection, and the result of the rendering function is the static (classic) visualization V_D. The control function $T_D(SF, Input)$, is implemented using control tools and helps to carry out the transfer of the search window in space, thereby changing the visualization. The presence of a control function in this model makes the visualization interactive. The control function and the visualization function consist of control and visualization sub-functions for each dimension of the space: D: $T_{D_1}(\langle p,w\rangle_{D_1}, Input)$ и $V_{D_1}(\langle p,w\rangle_{D_1})$.

$$D_i = \{o \in \langle s,p,o\rangle \in G \mid i = p\}$$

$$P_D = \langle V_D(SF), T_D(SF, Input)\rangle$$

$$V_D(SF) \to V_D$$

$$SF = \langle\langle p,w\rangle_{D_1}, \langle p,w\rangle_{D_2}, ...\langle p,w\rangle_{D_i}\rangle$$

$$SP \in SF$$

$$V_D(SF) = \langle V_{D_1}(\langle p,w\rangle_{D_1}), V_{D_2}(\langle p,w\rangle_{D_2}), \ldots V_{D_i}(...)\rangle$$

$$T_D(SF, Input) \to SF$$

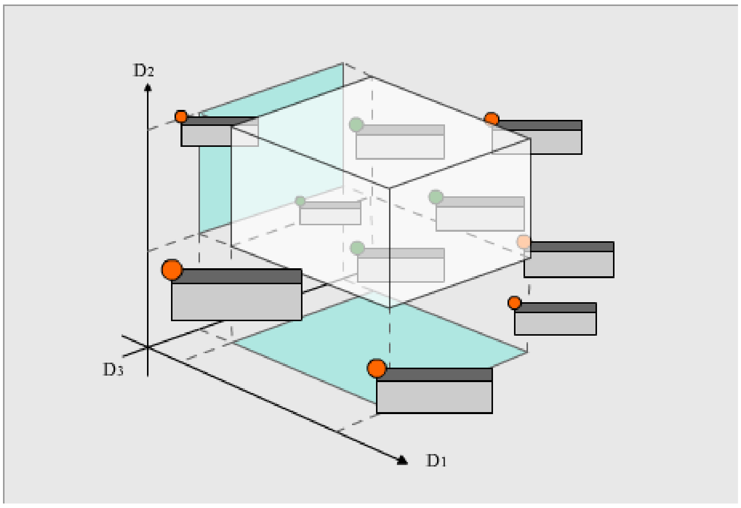

Рисунок 2 — 3D Data space

$$T_D(SF, Input) = \langle T_{D_1}(\langle p, w\rangle_{D_1}, Input), T_{D_2}(\langle p, w\rangle_{D_2}, Input), \ldots T_{D_i}(\langle p, w\rangle_{D_i}, Input) \rangle$$

Next, in the second chapter, there is a description of the task-oriented metrics of the quality of interactive visualizations. In the article «Quality metrics for information visualization», the authors propose to evaluate the quality of representation by three input parameters: Data, User, Task. Thus, in the article, voicing the fact that visualization of data cannot be of high quality if it does not allow solving the task initially set for it. Using this idea as a basis, the main scenarios for using visualizations of ontological data are highlighted. Most scenarios boil down to finding one, two or many elements of a view, so the search operation is taken as a basis. There are four types of steps in a search operation that can be arranged in a different order: «Input», «Visualization», «Overview», and «Perception». The simplest search scheme with a description of how the results of the dissertation are applied is shown in figure 3.

At the «Input» step, the user enters data, thereby moving the search box, the estimate of this step is calculated based on the distance the user has moved in space and how many simple operations he performed for this. At the step "Render"the rendering is calculated. «Visualization» can be assessed by metrics such as aesthetics and form metrics. «Overview» is the movement of the user in the data space within the search box. The user's position in the data space is determined by the search point, which, in particular, can be considered the point on the screen, at

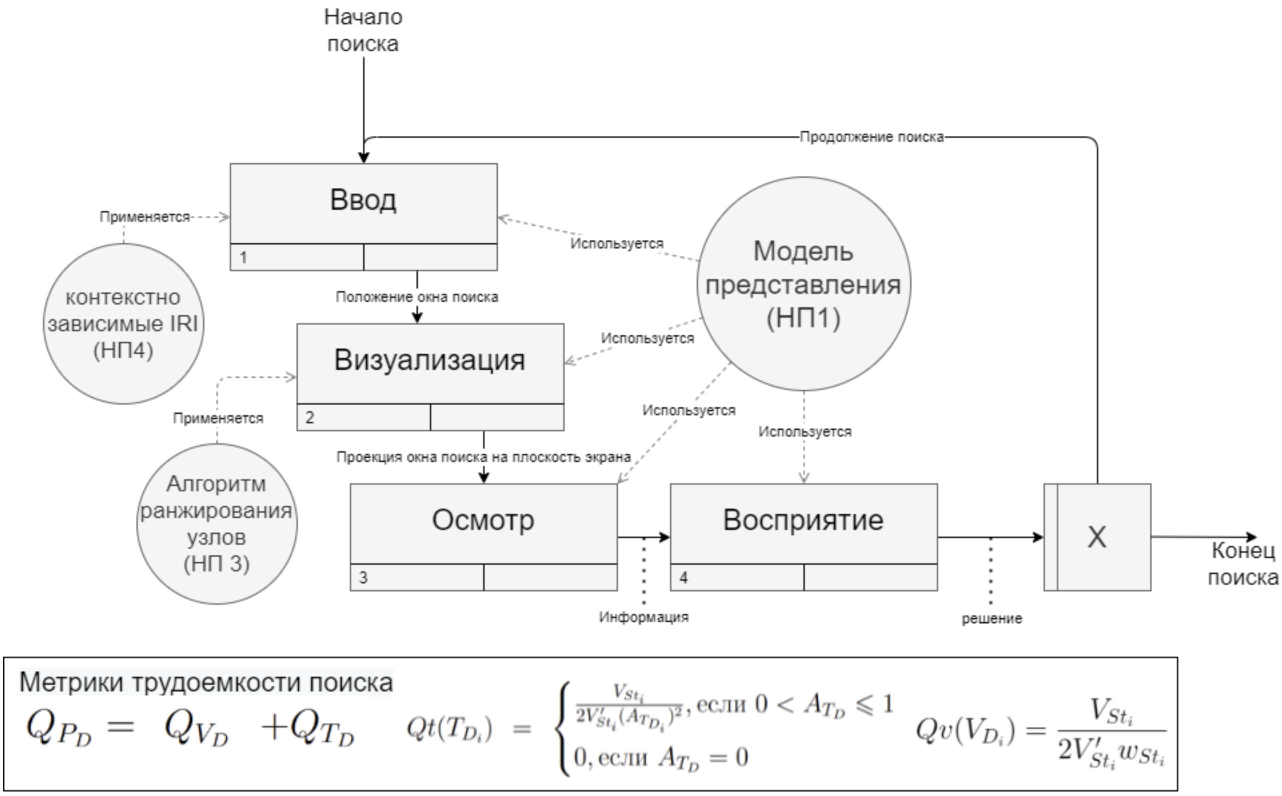

Рисунок 3 — Ontological data visualization method and search procedure

which the user's gaze is directed. Can be installed using EyeTracking technology. «Perception» is a step that exists to end a search or to decide on a new search. The search procedure is at the heart of task-oriented quality metrics or search complexity metrics. According to metrics, the quality of interactive visualization consists of two components: the quality of visualization (static) and the quality of management tools. Figure **??** shows the relationship between scientific results, interactive visualization model and search labor intensity metrics, as well as the main steps of the search procedure.

The quality of a control for interactive visualization *is the sum of the average value of the quality of each control tool, where the quality of each individual control is the ratio of the speed of the search step corresponding to this control to the ideal speed for all possible search operations, multiplied by the square of the space coverage. Where space coverage is the ratio of the volume of the available search space to the volume of the entire space, and the ideal speed is equal to the maximum distance between positions in the data space.*

$$Qt(T_D) = \frac{1}{n}\sum_{i=1}^{n} Qt(T_{D_i}) \mid$$
$$Qt(T_D) \in \mathbb{R}, \ \ 0 \leqslant Qt(T_D) \leqslant 0.5$$

$$Qt(T_{D_i}) = \begin{cases} \frac{V_{St_i}(A_{T_{D_i}})^2}{2V_{max}}, \text{если } 0 < A_{T_D} \leqslant 1 \\ 0, \text{если } A_{T_D} = 0 \end{cases},$$
$$St_i \in [Input],\ 0 \leqslant Qt(T_{D_i}) \leqslant 0.5$$

V_{max} – Ideal speed. $V_{max} = \lceil S_s \rceil \mid Qt(T_{D_i}) \in \mathbb{R},\ 0 \leqslant Qt(T_{D_i}) \leqslant 0.5$.

$A_{T_{D_i}} \mid A_{T_{D_i}} \in \mathbb{R},\ 0 \leqslant A_{T_{D_i}} \leqslant 1$ – The ratio of the existing data space to that available for the tool corresponding to the data dimension D.

Overall coverage ($A_{T_D} \mid A_{T_D} \in \mathbb{R},\ 0 \leqslant A_{T_D} \leqslant 1$) is equal to the ratio of the entire data space to the available.

The visualization quality is assessed in a different way. The rendering functions as part of the interactive rendering correspond to the steps of the search process of the type **Render**, but their quality is assessed by the steps of the **Inspect** type following the rendering step. **Render quality for presentation** *is the average of the quality of each aspect of the render, for all possible searches, where the quality of each individual render is half the ratio of the speed of the scan step following the render step to the ideal speed multiplied by the relative width of the search box. In this case, the ideal speed is equal to the maximum distance between positions in the data space, and the relative width of the search window is the value of the width of the search window for each search step, measured from zero to one relative to the maximum width of the search window*

$$Qv(V_D) = \frac{1}{n}\sum_{i=1}^{n} Qv(V_{D_i}) \mid Qv(V_D) \in \mathbb{R},$$
$$0 \leqslant Qv(V_D) \leqslant 0.5$$

$$Qv(V_{D_i}) = \frac{V_{St_i}}{2V_{max}w_{St_i}} \mid$$
$$St_i \in [Overview, Perception] \mid Qv(V_{D_i}) \in \mathbb{R},$$
$$0 \leqslant w_{St_i} \leqslant 1$$
$$0 \leqslant Qv(V_{D_i}) \leqslant 0.5$$

V_{max} – Ideal speed $V_{max} = \lceil S_s \rceil$.

The third chapter begins with a description of the essence of the dissertation work, that is, the method of visual management of ontological data in three-dimensional virtual space. In this work, the method is a technical way of developing tools for visual management of ontological data. The method is based on an interactive visualization model and is actively used at the stage of developing interfaces for visual ontological data management tools.

This is followed by a description of the prototype developed during the dissertation work, which allows you to test the method of visual management of ontological data formed during the work. The prototype is built on the existing Ontodia tool, so the rationale for choosing this tool as a base is first given. Then we discuss the structure of the basic Ontodia tool - its technology stack, architecture, data circulation paths. The Onodia architecture is shown in Figure 4. As you can see from the figure 4 the basic Ontodia tool consists of three main parts (Model, View and Controller), one of which (Controller) is split into two subparts (DiagramView and EditorController).

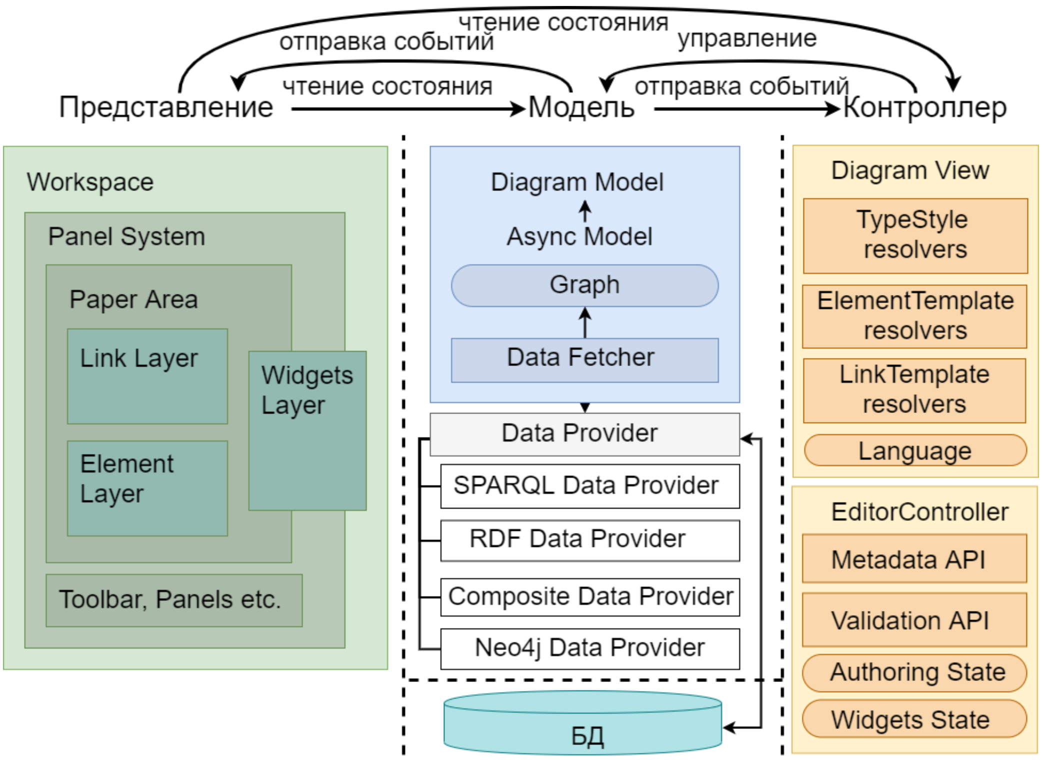

Рисунок 4 — Architecture of Ontodia

The following is the architecture of the developed prototype. The Ontodia3d architecture is shown in Figure 5. As you can see from the figure, it does not differ much from the basic architecture, but it extends it. The part of the *Workspace* object that is responsible for graph representations is duplicated. – There are two interchangeable implementations in the prototype, which makes it possible to switch from 2D to 3D. Also, the 3D view has a more complex internal architecture similar to that of the underlying tool. Before building the view, the underlying tool model is translated into the $L3 - Graph$ tool's internal model.

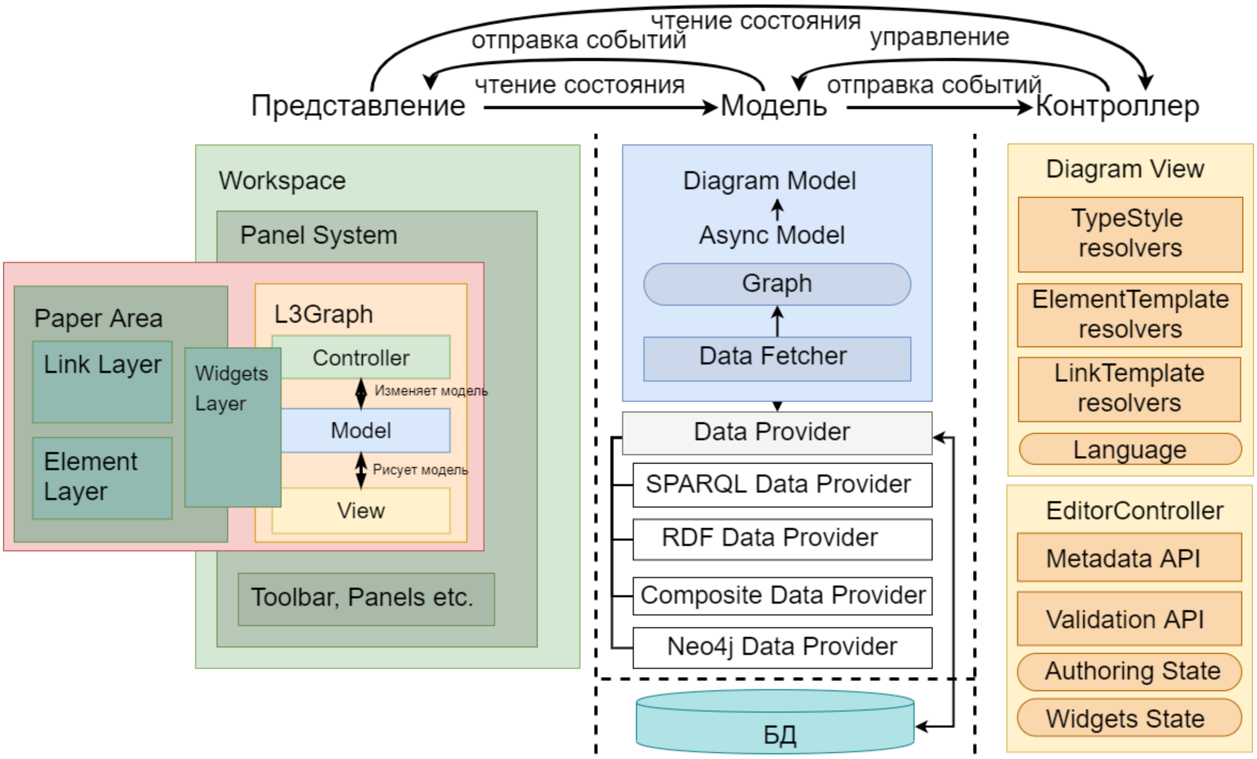

Рисунок 5 — Architecture of Ontodia3d

The second part of the third chapter describes an algorithm for generating context-sensitive identifiers for anonymous nodes, which allows to increase the coverage of the data space of the interactive visualization method. The main material on this section was presented in articles «Approach to Blank Node Processing in Incremental Data Visualization by the Example of Ontodia». The context of a BN (Blank node) is a subgraph of the main graph, which includes the target anonymous node, as well as transitively all anonymous nodes enclosed between the usual nodes surrounding this subgraph and including them (see fig. 6. That is the context for the target anonymous node is the graph that contains the target anonymous node and all of its neighbors, and the neighbors of its neighbors, up to the first regular node.

The solution we propose for displaying anonymous nodes is to introduce a preprocessing stage for queries and issue context sensitive identifiers to empty nodes. The preprocessing stage includes several steps: collecting the context, generating context-sensitive identifiers, saving the preprocessing results for subsequent issuance.

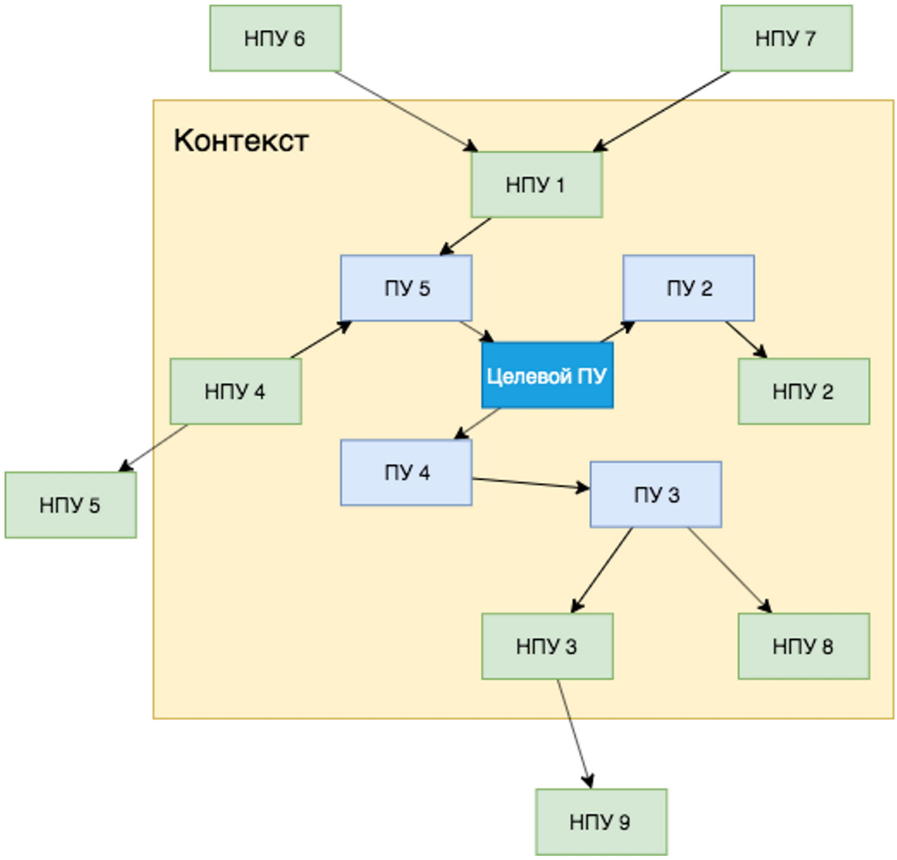

Рисунок 6 — Presentation of the BN context

The identifiers must be composed so that it is possible to reconstruct the context directly from each generated identifier and compare the nodes unambiguously. Briefly, the algorithm is as follows: a context graph is taken, then it is canonized (using an algorithm taken from the article «Canonical Forms for Isomorphic and Equivalent RDF Graphs: Algorithms for Leaning and Labeling Blank Nodes»). The identifiers of the canonicalized graph are encoded using the js-function encodeUriComponent, then the vocabulary of terms is extracted and compressed using the prefix tree. Next, the «dangerous characters» are replaced, and the indices of the elements are added. Now, firstly, it is possible to compare nodes, and secondly, to visualize them lazily. That is, decrypting the identifier, give out the following nodes.

Next, the last component of the method is described, namely, the node ranking algorithm using vectorization of the query and the original graph, as well as the attention mechanism for ontological graph data. The algorithm is based on the use of a neural network with the «Transformer» architecture to reduce the complexity of search steps such as «Overview». The attention mechanism is a kind of attention mechanism, the task of which is to identify patterns between input and output

Параметр	**Шаблон**
Example:The wine which is the instance of the Example:class Loire and which has property locatedIn and Example:located in Anjou Region is...	
Start	The whine which is...
typeInput	**Instance:** ...instance of the class {type}... **Class:** ...is a subclass of {type}...
propertiesInput	...and which has property {property}[and]...
subjectTypeInput	**Instance:** ...is the instance of the class... **Class:** ...is the class of wines... and is a subclass of Proposals connect through a union «and».
searchKey	and {searchKey}...

Таблица 1 — Natural language templates

data. During the search operation, the user queries the system using a graphical interface. The developed prototype generates a JSON-object, which using a set of templates (templates for substitution are given in the table 1). we translate into natural text and then send it to the GPT-2 language model where it is vectorized and processed. GPT-2 completes the text in the most natural way in accordance with its functionality, we break the completed text into terms and try to highlight the IRI using the lookup mechanism (Reconciliation API) implemented in the prototype. After that, we get a set of weights and rank the elements according to them.

The method of teaching a neural network is of scientific value here. GPT-2 by itself is capable of giving adequate answers to the questions posed, however, with additional training, the results are more accurate. As you know, ontological data are often represented as triplets <s, p, o> <subject, predicate, object>. Each of them usually has a name tag, or is itself a literal. In this case, predicates are usually represented in the form of verbs, so we can form natural language sentences practically by forming sentences from triples. The only problem is that not all predicates are presented in the verb form, for them we have manually entered the nominal labels (they are presented in the table 2).

Iri	Метка
$rdfs : label$	has label;
$rdf : type$	has type;
$rdfs : subclassOf$	is subclass of.
$owl : intersectionOf$	is intersection of
$owl : unionOf$	is union of
$owl : Restriction$	is restriction of

Таблица 2 — Additional predicate labels

Also, the neural network is able to remember the results of search procedures and try to take them into account during the next search. The last part of the chapter provides a set of recommendations for improving interactive visualization according to metrics of search effort at different steps. Recommendations are divided into two types: for the «Input» step and for the «Visualisation» step. This section contains descriptions of approaches to the visualization of diagram elements, methods of generating informative labels for elements and recommendations for the selection of layout algorithms.

The fourth chapter provides a detailed description of the experimental part of the study. The chapter begins with a description of a test bench built on the Metaphacts Platform. The Metaphacts Platform is a platform for creating WEB applications that manage knowledge graphs. The platform aims to support various categories of knowledge graph users by implementing appropriate services to manage knowledge graphs and by providing a rich and customizable user interface that enables the rapid creation of applications for specific use cases. The platform is built on open standards, which makes it reusable in different areas, and also simplifies the integration of knowledge graphs with different parts of organizational data and software infrastructures. Immediately after the description of the platform, the rationale for its choice and the ways of using its functionality are given. Then the software extensions for the platform, prepared in the course of the dissertation work, are described. Among the software extensions there are the following components: The component for calculating and visualizing the metrics of the search complexity and the Ontodia3D component that implements the described method. Then the search space, formed for the experiment, is described, which consists of such dimensions as: Three dimensions of space, Type of elements, Type of object (Instance or Class), Object properties and Attributes.

The next section of the chapter is devoted to the dataset that is used to conduct the experiment - the wine ontology (http://www.w3.org/TR/owl-guide/wine.rdf). It briefly describes the structure of the wine ontology and the main constituents. The following is a description of a method for generating test samples. Based on the description of the ontology, a standard list of questions was developed for various instances of the ontology. Test questions were composed as follows:

1. Choosing a copy of wine X;
2. We select two properties from the list [producer, grape varieties, region, color, sugar, consistency] ;
3. We give an assignment of the type: Construct a diagram that answers the question: «What value of properties 1 and 2 does the wine with the identifier X have?».

The next section of the chapter is devoted to the test procedure. Metrics are task-oriented and calculate the quality of interactive visualization based on the speed of the search procedure, respectively, to check the metrics and work at the stand, a list of search tasks was formulated that were solved by users based on the provided data set. The problems were formulated according to the principle described in the article «A diagrammatic approach for visual question answering over knowledge graphs» [0]. The subjects are asked a series of questions, the answer to which must be given using visualization. All participants were instructed on how to work with the developed prototype, and a short introduction to the field of ontologies, semantic technologies and graph theory was given. As a result, at the time of testing, the subjects showed, on average, the same result when working with the prototype, so it was decided to abandon the division of subjects into groups and conduct tests on one group of twenty-one people. The test procedure is followed by a section describing the bench control tools, in other words, the prototype control function.

The test bench is shown in Figure 7. In the figure 7, the letters mark the control tools corresponding to the previously defined dimensions of space. **a** is a type space management tool, it is a part of the filters panel, in the Metaphacts platform it is implemented based on the facet search component. **b** is a control tool in the object type space (class / instance), as well as the **a** tool is part of the filter panel. **c** is a property and attribute space control tool, as well as the **a** and **b** tools are part of the filter panel. **d** is a tool for managing three dimensions of space. This tool is built into the basic prototype and allows you to change the viewpoint to a 3D diagram.

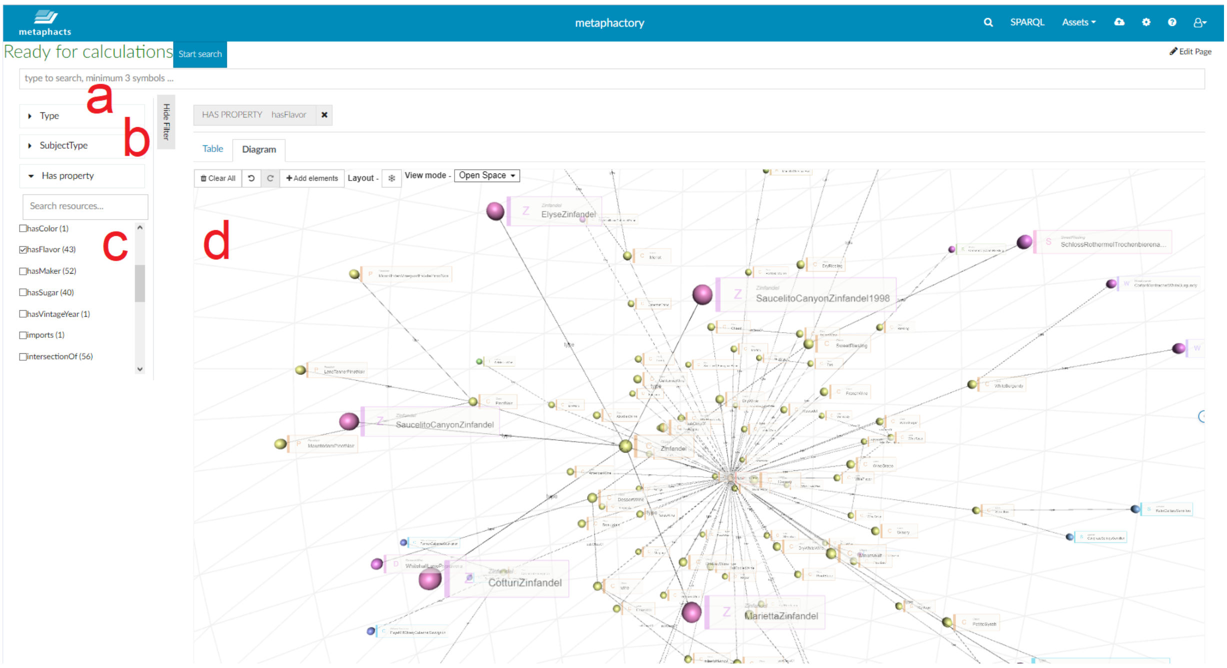

Рисунок 7 — Test bench: a – control tool in the type space; b – control tool in the object type space (instance class); c – control tool in the space of properties and attributes; d is a control tool for three dimensions of space.

– In the first implementation, it allowed the user to move along three axes of space, without the possibility of changing the viewing angle, the movement was carried out both using a computer mouse and a keyboard.

The next two sections of the fourth chapter are devoted to the results of the preliminary and final tests, after which there is a comparison of the results. The section provides test graphs and compares the test results of the basic tool and the tool that implements the proposed methods and algorithms. Figure 8 shows the average graph of approaching the target during the search operation.

The prototype was modified in two stages. First, a module was integrated into the prototype that made it possible to generate context-sensitive identifiers for anonymous nodes, then the interface was extended with a module that provides ranking of nodes using the mechanism of attention and vectorization of the original graph request. From the diagrams in Figures 9 and 10, we can see that the total number of switches between the «Input» and «Overview» steps has decreased by one switch. As you can see from the diagram in figure 11, the average labor intensity of input steps has decreased by 22% from the initial one, which indicates an increase in the quality of management tools. The user is less likely to use the Filter Pane, but they are increasingly using the Keyed Filter component, which provides intelligent filtering and maintains the ranking of items according to attention coefficient values.

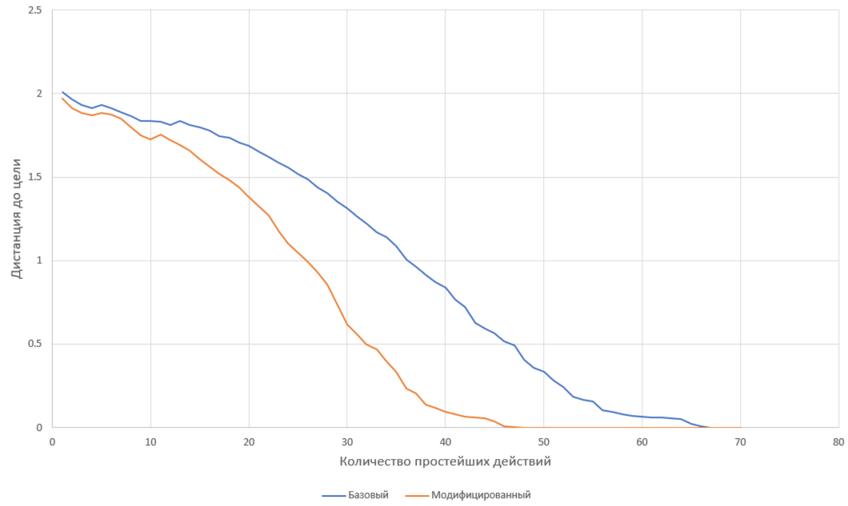

Рисунок 8 — Comparison of graphs of approaching the target

- This has reduced the overall labor intensity of the input steps. Adjusting the viewpoint is now also less time consuming, which is less time consuming. Comparison of graphs of approaching the target during the search process is presented in the diagram 8.

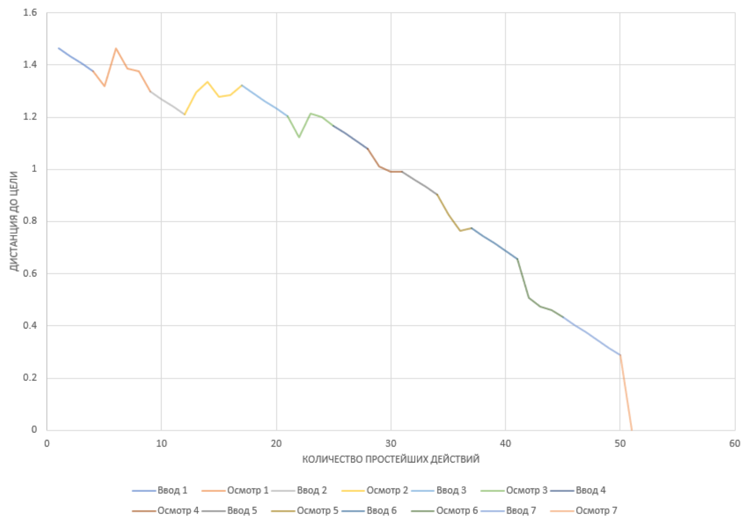

Рисунок 9 — Approximation plot using the example of one search procedure (basic version of the tool)

The initial coverage was 31% due to the inability to navigate to the anonymous sites. With the introduction of context-sensitive identifiers, the coverage reached its maximum value, which increased the value of control quality metrics. The diagram «Comparison of the results of calculating the metrics of the search complexity for different versions of the tool» (see Figure 12) shows the results of calculating the metrics of the quality of interactive visualization, as well as the quality of controls and visualization.

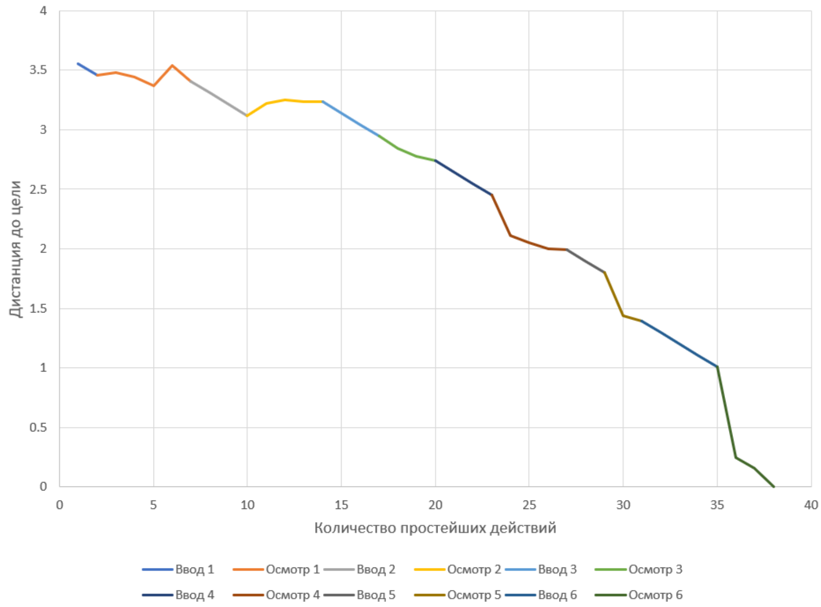

Рисунок 10 — Approximation plot using the example of one search procedure (modified version of the tool)

Moving on to the assessment of labor intensity at the steps of «Overview», it should be noted right away that the method for generating context-sensitive identifiers was developed for visualization within the framework of global visualization, therefore, it did not affect the quality of visualization in our case. At the same time, the addition of a module that provides ranking of visualization elements using the attention mechanism has significantly reduced the complexity of the «Overview» steps. After completing the «Input» steps, the target elements were visually closer to the user on the screen, so the user began the inspection procedure with them and quickly found the target element. The change in the average labor intensity when comparing the basic and modified versions of the tools is shown in the diagram in the figure 11. The decrease in labor intensity gave an increase in the quality of visualization, which, in turn, had a positive effect on the quality of interactive visualization (see the diagram in the figure 12).

The diagram in Figure 12 illustrates comparisons of rendering quality, tools and interactive rendering in general for the basic version of the tool, the version of the tool with the module for anonymous nodes enabled, and for the final version of the tool with the module for ranking nodes.

The final section of the chapter describes the practical value and use cases of the prototype and related techniques for enhancing interactive visualizations. The prototype is written in the form of a software module that can be embedded

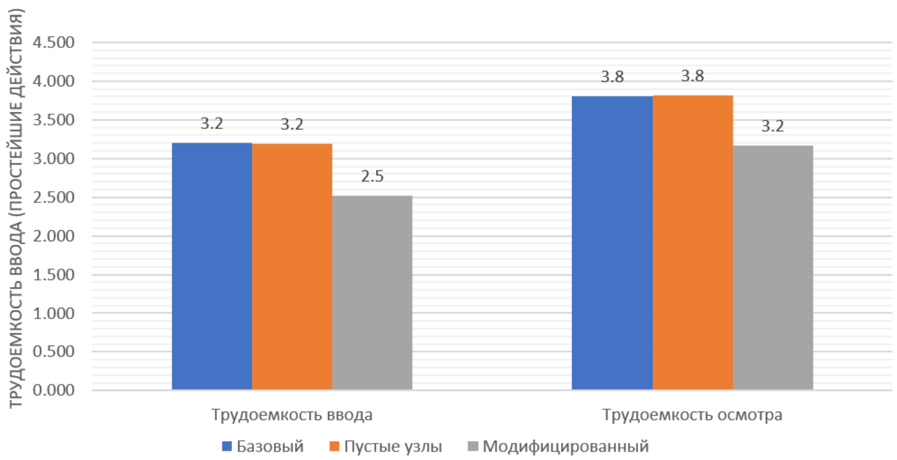

Рисунок 11 — Comparison of the average effort of searching for the input and inspection steps

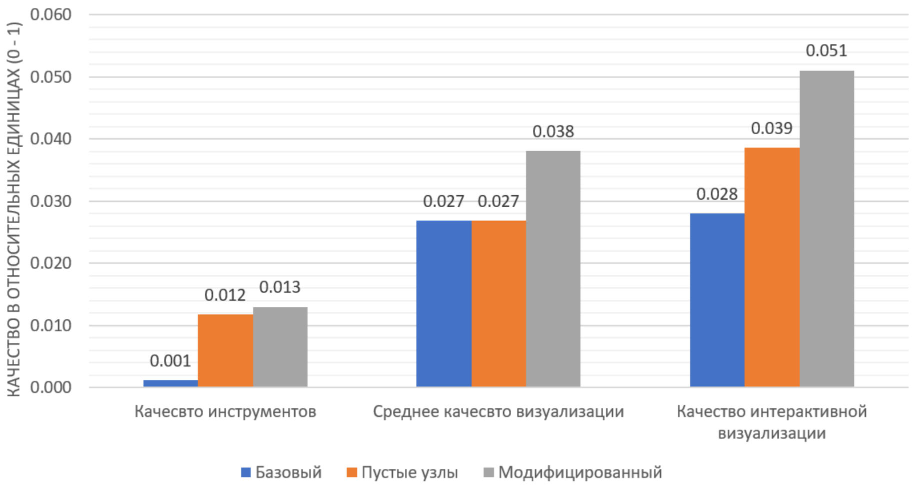

Рисунок 12 — Comparison of the results of calculating the search complexity metrics for different versions of the tool

in various web applications, however, the possibility of using the prototype in the projects of the British Museum is noted as the most useful use case. In the course of cooperation with the Museum in several related projects, the museum staff provided three-dimensional images of the terrain (see figure 13) models and ontological data.

An example of visualizing British Museum data is shown in figure 14. To generate a picture 14, the ability to customize the background image and define textures of three-dimensional objects has been added to the basic functionality of the prototype.

Another possible application turned out to be a scenario related to bioengineering data visualizations. *Wikidata* provides data on the structure of proteins and their interrelationships in an open ontological form. For example,

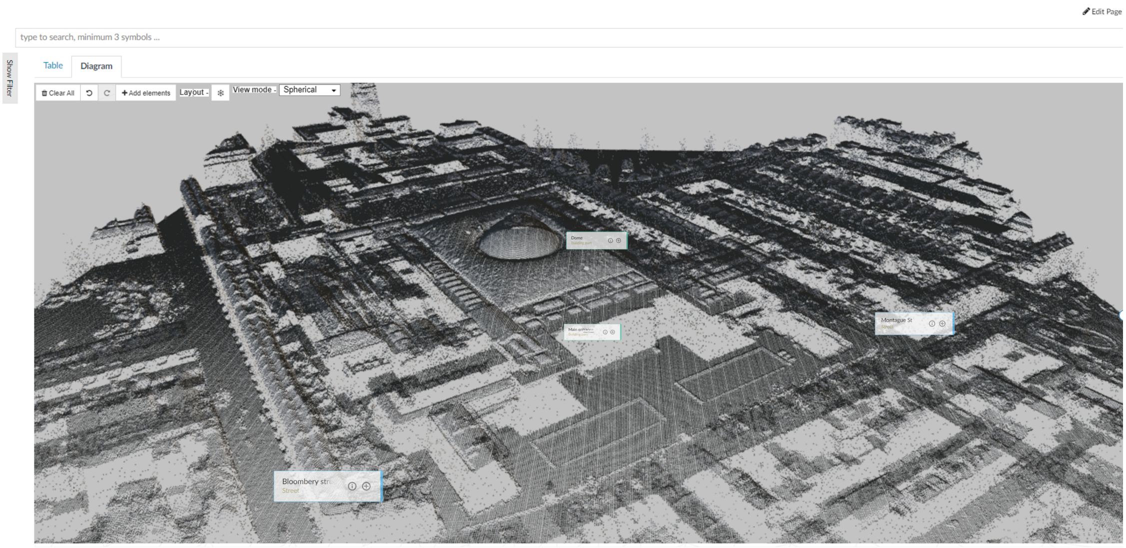

Рисунок 13 — LIDAR 3D image of the area around the British Museum

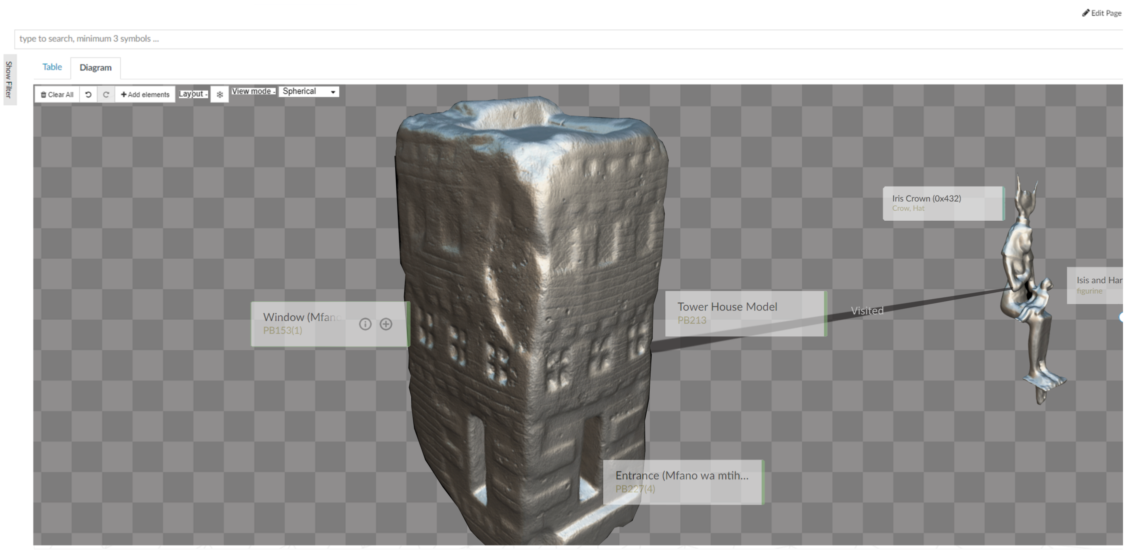

Рисунок 14 — Visualization of the 3D model of the Tower and its connection with the IRIs statuette

some pharmaceutical companies are interested in visual representation of transitive connections between mutations of certain proteins and diseases that these mutations cause. The *Metphactory* platform supports this use case and provides 2D visualization of bioengineering data, and 3D visualization of proteins (‘https://wikidata.metaphacts.com/resource/wd: Q21114990? View = page‘), however they are currently two separate use case (see fig. 15 A and B), which implies the loss of some context when switching between scenarios. The developed prototype allows combining the usage scenarios shown in the figure 15 and performing combined data visualization, supported by three-dimensional visualization of proteins (see figures 16 and 17).

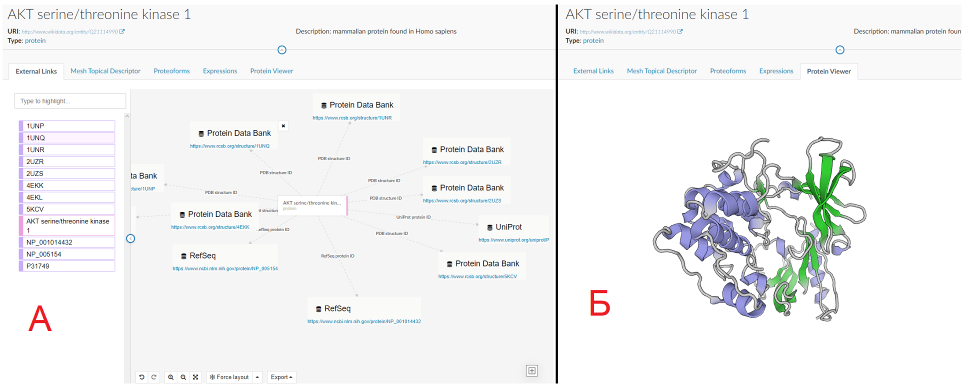

Рисунок 15 — 2D Bioengineering Data Visualization

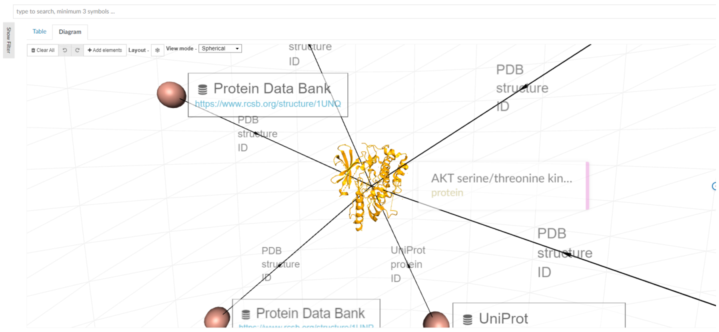

Рисунок 16 — 3D Bioengineering Data Visualization (1)

Conclusion

Within the framework of this work, solutions have been obtained that allow to reduce the time and labor intensity of interactive search in ontological databases using tools for constructing three-dimensional visualizations in virtual reality.

In the course of an analytical review of methods for representing ontological data in three-dimensional space, the limitations of existing solutions were identified and requirements for the construction of methods and means of visual control of ontological data in three-dimensional space were formulated, as well as a list of characteristics of visualization tools included in the prototype of the experimental stand.

The developed model of interactive visualization, which includes structural components and functions for managing ontological data in a multidimensional data space, made it possible to mathematically describe the methods for assessing three-dimensional interactive data visualizations, and the metrics developed on the basis of the model made it possible to objectively approach the issue of assessing the constructed visualizations and prove the consistency of the developed method. The visualization parameters were identified, which should be relied on in the process of making technical decisions.

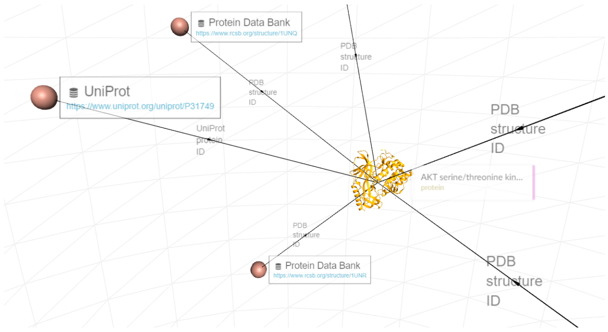

Рисунок 17 — 3D Bioengineering Data Visualization (2)

The available language models based on neural networks and the «Transformer» architecture have been investigated. Taking into account the requirements imposed by the process of forming a textual representation of ontological data, as well as the features of vectorization of semantic queries, the GPT-2 language model was chosen. On the basis of this model, an algorithm for ranking the vectorized nodes of the ontological graph was developed, which made it possible to reduce the complexity of searching in the data space by reducing the number of relevant simplest operations.

The problem of incomplete coverage of the space of ontological data in the process of lazy visualization of fragments of ontologies containing anonymous nodes is solved. The solution is based on the formation of context-sensitive identifiers of anonymous nodes, which allow restoring the canonicalized context of these nodes and continuing to navigate to ontology zones previously inaccessible for search.

An experiment was planned and a prototype visualization tool was developed for the experimental bench on which the tests were carried out. Tests have shown an improvement in the values of the search complexity metrics for interactive visualizations, which objectively showed the efficiency of the proposed approach.

The calculation of the search complexity metrics for the basic prototype made it possible to identify bottlenecks and identify ways to improve interactive visualization. So the first bottleneck was the coverage of data space with search tools. It turned out that the initial coverage was 31%, – such a low value was obtained due to the inaccessibility of search tools for data space zones containing anonymous nodes.

The second bottleneck is the complexity of the process of managing the constructed visualization. To reduce the complexity of the search, a set of recommendations for the construction of visualization was developed, as well as an algorithm for ranking nodes, which uses language models and the mechanism of attention. The results of the experiments showed a decrease in the complexity of the search at various stages from 36% to 46%.

The implemented software library for visualization methods using the algorithm for issuing context-sensitive identifiers and the attention mechanism for ordering nodes was used in a number of scenarios for working with real ontological knowledge bases and three-dimensional graphic models.

Evaluation of scenarios for using the practical component of the work demonstrated the possibility of using the prototype both in the tasks of visualizing data presented in the Web Ontology Language, and for representing connected data in general. In particular, the application of the results to research data on the cultural heritage of the British Museum, as well as to work with ontological data in the field of bioengineering, is demonstrated.

Thus, the goal of the work has been achieved, and all the tasks set within the framework of this dissertation research have been completed.

List of publications on the topic of the dissertation

1. Раздьяконов Д.С., Морозов А.В., Павлов Д.С., Муромцев Д.И. Подход к обработке пустых узлов при порционной визуализации данных на примере инструмента ONTODIA // Программирование - 2020. - № 6. - С. 16-29
2. Razd'Yakonov D., Morozov A., Pavlov D., Muromtsev D. Approach to Blank Node Processing in Incremental Data Visualization by the Example of Ontodia // Programming and Computer Software - 2020, Vol. 46, No. 6, pp. 384-396
3. Razdyakonov D., Wohlgenannt G., Emelyanov Y., Pavlov D., Mouromtsev D. A New Tool for Linked Data Visualization and Exploration in 3D/VR Space // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2019, Vol. 11762, pp. 167-171
4. Razdyakonov D. S., Muromcev D. I. Application of an Agent-Based Model to Develop Ontological Data Visual Management Tool (2023). Scientific Visualization 15.5: 73 - 88, DOI: 10.26583/sv.15.5.07
5. Раздьяконов, Д. С. "МЕТОД РАЗРАБОТКИ СРЕДСТВ УПРАВЛЕНИЯ ОНТОЛОГИЧЕСКИМИ ДАННЫМИ В ТРЕХМЕРНОМ ВИРТУАЛЬНОМ ПРОСТРАНСТВЕ."Актуальные вопросы современной науки: теория, технология, методология и практика. 2023.
6. Wohlgenannt G., Klimov N., Mouromtsev D.I., Razdyakonov D., Pavlov D., Emelyanov Y. Using word embeddings for visual data exploration with ontodia and wikidata // CEUR Workshop Proceedings - 2017, Vol. 1932
7. Wohlgenannt G., Klimov N., Mouromtsev D., Razdyakonov D., Pavlov D., Emelyanov Y. Using word embeddings for search in linked data with Ontodia // CEUR Workshop Proceedings - 2017, Vol. 1963, pp. 16-24
8. Mouromtsev D., Pavlov D., Emelyanov Y., Morozov A., Razdyakonov D., Parkhimovich O. Workflow supporting toolset for diagram-based collaborative ontology development implemented in the open budget domain // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2016, Vol. 9989, pp. 178-182

9. Mouromtsev D., Pavlov D., Emelyanov Y., Morozov A., Razdyakonov D., Galkin M. The simple, web-based tool for visualization and sharing of semantic data and ontologies // CEUR Workshop Proceedings - 2015, Vol. 1486, pp. 77