TODOS as digital science-support environment to provide STEM-education

Yevhenii B. Shapovalov\textsuperscript{1}[0000-0003-3732-9486], Viktor B. Shapovalov\textsuperscript{1}[0000-0001-6315-649X] and Vladimir I. Zaselskiy\textsuperscript{2}

\textsuperscript{1} National Center “Junior Academy of Sciences of Ukraine”, 38/44, Dehtiarivska Str., Kyiv, 04119, Ukraine
sjb@man.gov.ua

\textsuperscript{2} Kryvyi Rih Metallurgical Institute of the National Metallurgical Academy of Ukraine, 5, Stepana Tilhy Str., Kryvyi Rih, 50006, Ukraine
zaselskiy52@mail.ru

Abstract. The amount of scientific information has been growing exponentially. It became more complicated to process and systemize this amount of unstructured data. The approach to systematization of scientific information based on the ontological IT platform Transdisciplinary Ontological Dialogs of Object-Oriented Systems (TODOS) has many benefits. It has been proposed to select semantic characteristics of each work for their further introduction into the IT platform TODOS. An ontological graph with a ranking function for previous scientific research and for a system of selection of journals has been worked out. These systems provide high performance of information management of scientific information.

Keywords: TODOS, science environment, educational environment, ontology, taxonomy, STEM-education.

1 Introduction

1.1 The problem of digital science

Nowadays, cooperation and all-world international integration are conducted. Therefore, it leads us to the generation of a huge amount of not-structured information. One of the humanity actions fields which is one of the leaders of information production is science. The situation is being complicated due to the fact that providing science is foresees knowing of the huge amount of already made scientific researches.

Therefore, in science nowadays is a lot of information generated and there is a problem to process it. Considering this, educational approaches are adopting and one of the modern approaches which include principles of multidisciplinary and studying to work with a huge amount of knowledge is STEM-education. The specifics of it is the lack of digital instruments to provide it [2; 3; 20].
1.2 Scientometric databases in post-soviet countries challenge

For post-soviet countries, this situation is even more important due to the fast speed of integration of their science with worldwide which wasn’t provided previously.

Nowadays, in the example of Ukraine, the huge challenge is to provide publication in both well-known scientometric databases (such as Scopus or Web of Sciences) and journals recommended by Ministry of Education and Science of Ukraine (scientific professional editions of Ukraine; further – SPE). This makes informational chaos in the field of journal selection to publish the results of scientific work.

1.3 Information processing problem

As was noted before, a huge amount of scientific information is generated nowadays. However, there is no effective way to process them. Sure, systems which can simplify exists, such as Mendeley, but they still do not provide analysis and processing of the information. For example, well-known designs can only provide commenting of the scientific papers which isn’t provide any analysis and actually do not provide any systemizing of the information to provide structuration. The interface of commenting in Mendeley is shown in Figure 1.

![Fig. 1. The interface of commenting in Mendeley](image)

We propose using of Transdisciplinary Ontological Dialogues Object-Oriented Systems (TODOS) [6; 9; 13; 25; 26] to provide systemizing and processing of Big Data with taxonomy creation, filtering, and ranking of information. A key benefit of this system is the context-based method of data processing and structuring based on semantic relations. Previously, there was provided attempts to use ontology-based approaches in education [1; 7; 8]. However, they were characterized by not attractive for students and teachers’ interface and by low interactivity such as the absence of multiagency approaches or wasn’t interactive at all. In the case of Ukraine, it is very important to provide education in the national language and IT platform TODOS can implement it.
Therefore, this work aims justification of necessarily of digital science supporting processing and creation of the base of it.

2 Literature review and problem statement

2.1 TODOS as ontology systemizing of information

Using the ontological approach to provide informational systemizing is an important part of the learning process [4; 5]. Such an educational environment based on the ontological approach involves filling adaptive educational services with information resources that reflect the conceptual system of a particular discipline. The methodical provision of the educational-cognitive process consists of the assimilation of the conceptual system, axiomatics, rules, syntactic and morphological foundations of this theory. The set of terms determines the conceptual basis of scientific theories by determining a certain ordering of the concepts of the discipline. Thus, the ontological multiagent in content reflects the conceptual system of a certain disciplinary theory. It takes into account the individual characteristics of each subject of the educational process.

Structures in TODOS are represented by three categories (O, A, R), where O and A are a set of elements called objects and attributes, and R, respectively, is the binary relation between O and A. In particular, if oRa for o O, A, then we assume that “the object possesses the attribute A” or “the object has the attribute O” [7].

The feature of the ontological graph is the high level of structuring and data visualization, the possibility of transition between related edges and search for semantic links between vertices and its elements. The graph provides a transition to scientific data carried out quickly and understandably. In addition, operationality of information can be significantly improved by transforming it to taxonomy under using of ontological approach [6; 22].

TODOS is an innovative complex of programmatic information and methodological knowledge management tools using ontological management approaches to corporate information resources, where people are considered as the source of the birth of new knowledge for transferring them in the form of their own knowledge through the tool TODOS, which is the only integrated point of access – “the single window” – to the information and applications of the system to provide interactive interaction with users. A key benefit of this system is the context-based method of data processing and structuring based on semantic relations.

The architecture of the formation of transdisciplinary information environments IT-platform based on the multiple procedures of transdisciplinary interaction with network information resources is realized on the basis of semantic control and ontological interface of TODOS [23; 26]. The technical basis of the TODOS is consisted of [25]:

— SYSTEM CONSPECT provides the construction of terminological trees based on the analysis of natural language text. It is a linguistic processor that provides the initial formation of a linguistic case and allows to solve the following practical tasks: improving the quality of processing of linguistic texts by increasing the vocabulary
of the system; automatic definition of thematic directions of the document; sorting of documents according to thematic directions.

- SYSTEM CONFOR provides the creation of ontology subject areas, classification, and generation of taxonomies in the form of ontological graphs. The system ensures the creation of subject area ontology, classification, and generation of taxonomies in the form of ontological graphs, which allows solving the following practical tasks: construction of a semantic network of terms of the document; combining semantic network of terms for several documents.

- SYSTEM EDITOR ensures the formation of ontological models through the creation, editing, review and analysis of networks of concepts based on the construction of semantic links between objects of the subject area and the formation of patterns, presented in the form of a set of values of attributes, which describes the initial concepts of subject areas. The isolation of regularities is carried out by the method of inductive formation of concepts based on the pyramidal network.

- ALTERNATIVE SYSTEM ensures the organization of objects-concepts of ontology, on the basis of integrated processing of properties that characterize them. For this, we use weight, ball, and linguistic scales. Each such scale defines the values of the criteria characterizing the properties of the objects of the thematic ontology of the subject area. In general, the properties-criteria are characterized by different degrees of importance, which when solving the problem of choice are given by some real numbers – weight coefficients. Before solving the problem for each criterion, it is necessary to form its value for each alternative. Thus, the formation of ontologies of the tasks of choice is ensured.

- LINGUISTIC CORPUS and built into its environment SEARCH MACHINE provides marking and indexing of semantic units that define and describe the contexts of objects of thematic ontologies of the subject areas. Contexts of semantic units make up an electronic library with means of associative search of semantically related information arrays, including determining the level of semantic equivalence of texts.

Fig. 2. Information management system TODOS
These modules are working together to transform unstructured incoming data to the hierarchy of contexts. Information management system TODOS is shown in Figure 2.

2.2 Main features of the TODOS: taxonomy, filtering, ranging

Ontologies are based on taxonomy creation. The main feature of the TODOS platform is a simplification of its creation. To create the ontology user do not need to know any programming languages just MS Excel. The example of taxonomy created on TODOS platform is presented in Figure 3.

Fig. 3. The example of a taxonomy provided by TODOS platform

Fig. 4. The example of an objective view of the ontology created in TODOS platform
To provide visualizing of the taxonomy, it’s possible using the objective view. This view presents each edge of the ontology as a personal object. The hierarchy is saved by creating links between those objects. The example of a taxonomy provided by TODOS platform is presented in Figure 4.

Analysis of information is provided through the identification and separation of semantic information of each edge. As edge, we can use any object kind scientific paper, single microorganism, the technology of water clearing, etc. It depends on the expert-creator idea, but anyway, it provides separation of the semantic data of edges. This can provide further data processing and systemize in way of filtering of ranking. The proposed informational system is characterized by multiagent features and has all the benefits of such a system.

3 Materials and methods

For creating digital educational programs and other educational content, the sheets were loaded to the part of TODOS IT-platform editor. After that, the generation of the graph edges with its characteristics was carried out.

To store information and provide its sharing, Google sheets were used, with their further conversion into the .xls and .csv MS Excel sheets (see in Figure 5).

The obtained documents were used to create the ontology structure .xml and to fill the ontology graphs with semantic and numeric information for ranking and filtering. Some of the instruments of the web-oriented educational environment are using intellectual features of TODOS, and to provide it, semantic characteristics were added.
The received documents were used to create an ontology structure (xls) and to fill the ontology graphs of ranking and filtering. To provide it, they were downloaded in editor4, the part of TODOS IT-platform. After that, the graph generation and the inputting of semantic characteristics to each vertex were carried out. Ontological edges were formed using predicate equations [25]:

\[
Pr(x_1, \ldots, x_n) = \begin{cases} 
1, & \neg Pr(YGr x_i) \land Pr(x_1, \ldots, x_n) \\
0, & Pr(YGr x_i)
\end{cases}
\]

where \( x_i \in X; 1 \leq i \leq n \).

The relation between taxonomic categories has the properties of the hyperrelation \( Gr \) type \( YGrx \), where \( Y \) is the set of all possible sets of concepts of \( X \) taxonomic category \( T \), \( x \) is one of the concepts of this set and \( Pr \) – predicate.

The obtained ontological graphs were opened in the appropriate form, ranking or filtering. To provide filtering, the function of choice has been applied. The function of choice in terms of taxonomic categories is as follows:

\[
\forall T[\emptyset \notin T \Rightarrow \exists F: T \rightarrow \cup \forall T \in \tilde{T}(F(T) \in T)]
\]

where \( F \) – is a function of the interpretation of a certain ontology; \( T \) – taxonomy.

4 TODOS as the digital science-support environment

All advantages of TODOS can be used to both systemize the science information and to create useful databases (Big Data based) instruments for the scientist.

4.1 Using TODOS to create Big Data databases

SPE and SCOPUS ontology-based selection systems.
We created the online web-oriented ontological graph for both, SPE and SCOPUS journals to provide selection. As graph edge, each journal was chosen. For both, semantic characteristics were separated. For SPE journals they were “Founder”, “Branch of science”, “Date of inclusion/renewal”, “Journal indexing”, “Journal specialization ». User can use those characteristics to select a journal due to its needs.

SPE journal selection instrument is presented in Figure 6.

To create a database on SCOPUS journals “SJR”, “SNIP”, “CiteScore”, “Activity status” (active or not), “All Science Classification Codes (ASJC)”, “Language in the source (three-letter ISO language codes)” and “Publisher’s Country” were separated from each journal and added to edges as semantic data. Scopus journal selection instrument is presented in Figure 7.

Ontology-based catalog of the microorganisms.
Systematization of knowledge in the field of biotechnology may also be complicated by the fact that semantic characteristics cannot always be quantified, and therefore the ranking system cannot always solve the issue of information management. For such
systems, it was suggested to separate the semantic information and apply a filtering function. The semantic characteristics of each microorganism were also proposed and input into the Google Sheets. All semantic characteristics were added in the collective access mode [19].

![Fig. 6. SPE journal selection instrument](image)

![Fig. 7. Scopus journal selection instrument](image)

The resulting ontological graph provides the possibility to use the filtering, and it is possible to find the discovered microorganism or group of microorganisms. General view of the ontological taxonomy of microorganisms is presented in Figure 8 and a general view of the microorganisms selecting system is presented in Figure 9.
4.2 TODOs to systemize scientific information

To construct a system of ranking of previous studies, we have identified semantic characteristics of the scientific research devoted to biogas production from chicken manure. These semantic characteristics include “Temperature (°C)”, “Volume of reactor (l)”, “Chicken manure content (%), “Moisture content (%), “Active sludge content (%), “Final solids content (%), “Biogas production (ml/g VS)”, “Methane production (ml/g VS)”, “Methane content (%), “Year of the research”, “Ammonium nitrogen content (mg/l), “Final pH”, “Initial pH”, “Minimal pH” and “Maximum pH”
The characteristics were selected from the studies on dry fermentation of chicken manure and were input to the google sheets.

The data were processed by the methods described in detail in our previous works [3; 16]. As a result, it was possible to use ranging from previous research results. The general view of the taxonomy is presented in Figure 10. The interface for selecting the importance of indicators is presented in Figure 11, and the interface for ranking the results is presented in Figure 12. The interface for selecting the priorities of numerical information for ranking allows taking into account the priority of modern articles, with the correct marking of important criteria. The considered system allows a quick search of the information by the necessary criterion [19].

![Figure 10. The general view of the taxonomy](image)

4.3 Transdisciplinary using scientific results in education and science.

Single science digital environment to provide STEM-education

As it was proposed previously, the ontology-based system can be used to provide integration and transdisciplinary using internal sources [20; 15]. Databases created by a group of scientists who provides research will be able to share it to the open-source general database. That knowledge can be used by a huge amount of people not just scientist. As it was proposed previously, the ontology-based system can be used to provide integration and transdisciplinary using internal sources. It means, that multidisciplinary ontology-based educational environment can’t be used as the main instrument which provides scientific method of education and can integrate other instruments of STEM-education such as augmented reality or mobile phones involving [10; 11; 12; 15; 20; 24]. The proposed system will be very useful for students and young scientists who just start their research work.
5 Conclusions

1. A huge amount of scientific information can be systemized by using TODOS IT-platform.
2. TODOS IT-platform can provide a high level of informational structuring and information processing through the creation of the hierarchy and using TODOS instruments such as ranking and filtering.
3. TODOS can be used to both systemize the science information and to create useful databases (Big Data based) instruments for the scientist.
4. We developed the method of systemizing scientific information which is characterized by a higher level of informational processing.
5. TODOS integrate the scientific processed data in a single scientific informational field which involves scientists and students to provide transdisciplinary researches.
6. The proposed system can be used not just for a huge amount of people not just scientist and provides integration of internal and external sources to provide research approach in STEM-education.

References


