Structurization and Processing of the Scientific Studies in the Form of Digital Ontologies

Yevhenii Shapovalov[1][a], Viktor Shapovalov[1][b], Roman Tarasenko[1][c], Stanislav Usenko[1][d], Adrian Paschke[2][e] and Iryna Savchenko[1][f]

[1]*The National Center "Junior Academy of Sciences of Ukraine", 38-44 Degtyarivska Str., Kyiv, 04119, Ukraine*
[2]*Fraunhofer FOKUS (with support of BMBF "Qurator" 03WKDA1F), Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany*

Keywords: Cognitive IT-Platform Polyhedron, Ontology, Ontological Tool, Scientific Studies, Scientific Reports, Learning Environments.

Abstract: Nowadays, there is a wide variety of scientific articles. Due to this fact, it is hard to read and be familiar with all of them. Also, it is hard for a young scientist to understand the complicated terms and methods used in a specific research domain. This problem was partially solved by bibliographic management software and other specific software. This article aims to develop an approach for structuration and processing sets of studies using the IT Platform Polyhedron using an ontology-based hierarchical model. The ontological graph is complex because it has additional branches from child nodes in its structure. The basis of our solution was IMRAD (Introductions, Methods, Results, Abstract, Discussion structure) which has been represented in the view of nodes. Those nodes have been connected with specific representations of IMRAD elements. Specific articles have been represented in the view of leaf nodes. That could help to use the taxonomies for the structuration of the articles. Each data block is in the form of separate attributes of the ontological node. The proposed solution allows to obtain structured sets of studies and separate their characteristics. Thus, the proposed ontology allows viewing all methods, measured parameters, etc., of the studies in a graph node structure and using them to find the used studies. The usage workflows to demonstrate mechanisms of the system's usage are presented. The method of merging a few graphs of studies is developed and presented.

1 INTRODUCTION

Usage of information technologies (IT) in various fields of research activities and the capability of software support in science to automatically classify and structure information, e.g., in publication data, becomes increasingly important. Nowadays, vast amounts of research data are available that is not structured, e.g., publications, presentations, etc. It is complicated for young researchers and scientists to use such large amounts of publication data. During the research process, young scientists are looking for, e.g., examples of research methods and parameters. However, this task is challenging at the early stage

of their scientific career. When preparing papers and reports, such literature search and analysis problems (e.g. state of the art) are challenging for every scientist (including youth and school researchers). For instance, according to Lens.org, the number of articles on biogas in 2002 was approximately 134, then in 2014, the number grew to almost 1164, as shown in (figure 1).

So, it is relevant to provide a solution that can simplify processing and information/knowledge extraction in scientific publications. There are two hypotheses in our study. The first one is about structuring and digitalizing the data, which can simplify finding the details about the research method. The second one is about structuring the results of previous studies, which can be represented as data of the informational system. Previously, this goal was partially achieved using metadata for data processing. In this paper, we further contribute with a semantic ontology and a more expressive semantic metadata approach.

[a] https://orcid.org/0000-0003-3732-9486
[b] https://orcid.org/0000-0001-6315-649X
[c] https://orcid.org/0000-0001-5834-5069
[d] https://orcid.org/0000-0002-0440-928X
[e] https://orcid.org/0000-0003-3156-9040
[f] https://orcid.org/0000-0002-0273-9496

Figure 1: Dynamic of the number of papers on biogas.

1.1 Using Metadata to Provide Data Management in Science Publications

It is relevant to use metadata about each paper to support publication data management. This metadata represents the data about the publication. In this case, the metadata can represent relevant information about each specific publication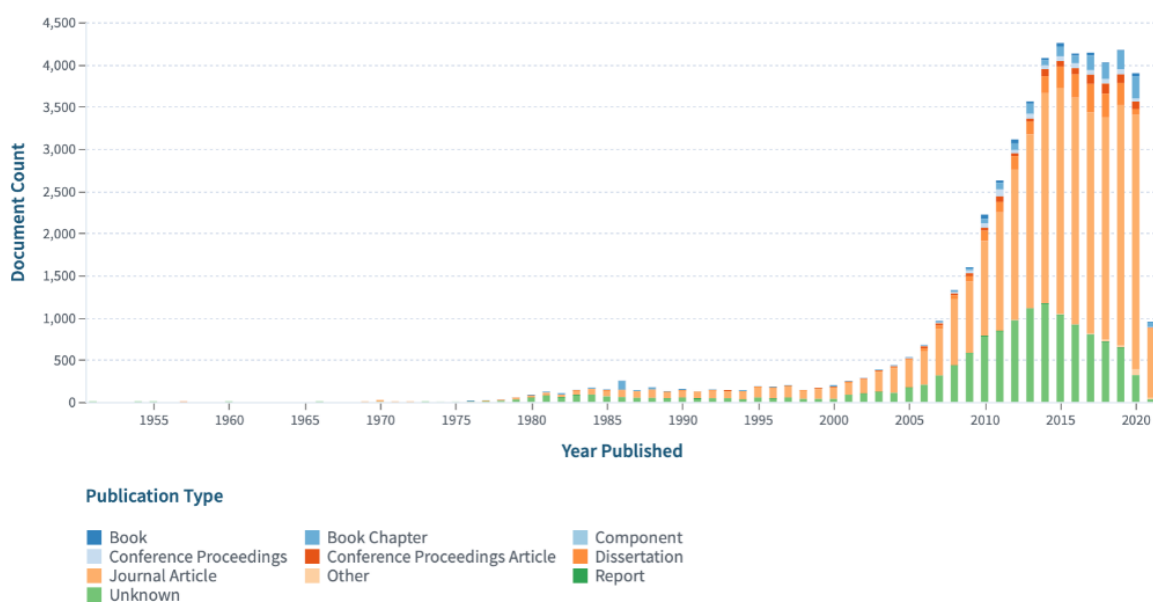. Metadata can include, e.g., contact information, year of publication, author details, instrument and protocol information, survey tool details and much more (Noyer et al., 2009).

For instance, reference management software maintains a database of articles and creates bibliographies and reference lists for the written works. This software simplifies the record of metadata. Several popular reference management software, for example, Refworks, Mendeley, EndNote and Zotero are used worldwide (Kumar Basak, 2014; Parabhoi et al., 2017; Salem and Fehrmann, 2013; Ram and Paul Anbu K., 2014). These managers can save profiles, build a database of citations, save PDF files, extract metadata, and import references from library catalogues, websites, and other citation managers (Ivey and Crum, 2018; Butros and Taylor, 2011; Cuschieri et al., 2019).

However, these systems only use limited metadata vocabularies without expressive semantic models. For instance, such systems do not support metadata concepts such as "Results," "Materials and methods," "References", etc. As a result, all these systems do not provide a systematic approach, and they are not

entirely semantically structured and not hierarchical.

1.2 Using Metadata Information of Scientific for Automatic Literature Review Processing

There are different types of metadata information that we can use to structure the articles, for example, by the relation to source, function, purpose, language, and publication date.

With using the a "Relation to Source" metadata the user defines the type of text that can be included in a classification program. A classification composed of extracts having exact sentences of a source document is known as an extractive summary. That is the simplest type of classifier.

With using a "Function" metadata the user can use any helpful and relevant information from source documents, for example, an abstract of a scientific article or the reviewers' opinion on the quality of work.

With using a "Purpose" metadata structures the article's purpose or main idea. The user needs to write down the general purpose or sense of the text in the program by himself.

With using a "Language" metadata, a classifier can be monolingual or multilingual is used. The monolingual classifier uses only one language and produces an output classification in the same language as the input document. In contrast, the multilingual classifier uses multiple languages and gives an output classification in one of the languages from the input document.

With using "Publication date" metadata, it is possible to arrange the articles by the time of their publication. To do so, the user must enter the article's publication date in the system, and then the system will arrange the article in an appropriate section.

There is a lack of methods for the structuration of scientific articles. Also, it is necessary to add metadata "By the Results" (each specific results that has been obtained during study) and "By the methods" methods (each specific method that has been obtained during study).

"By the Results" this method structures the article by its results. It provides possibility to process new numeric and semantic data that was obtained during research.

"By the Methods" this metadata is used structure studies by the scientific methods and matherials used in the study.

Considering that most articles have a typical IMRAD (Introductions, Methods, Results, Abstract, Discussion) structure, it seems advisable to build an algorithm that uses data from many specific articles to create ontological graphs that can be integrated with specialized educational institutions environments for young scientists.

1.3 Instruments for the Creation of Ontology-Based Learning Environments

A learning environment is a diverse platform where users engage and interact with learning new skills. While learners can learn in various settings, the term typically refers to a digital alternative to the traditional classroom. To improve learning efficiency and adaptability, formalized information resources that provide a high degree of structuring should be used in learning. An ontological approach could support this. The ontological approach provides a holistic and systematic approach to the study of various information sources and a specific subject domain, ensures the conceptualization and taxonomization of terms within the subject area and the existence of relationships between the terms of different subject areas to ensure multidisciplinarity. Computer ontologies are one of the effective mechanisms for ensuring a stable digital learning environment.

In recent years significant progress was made in developing ontologies. In this article, an "ontology" is a term that means a software or web system that consists of nodes with data. All ontology nodes are arranged in a specific hierarchical order, often referred to as an ontological tree or ontological graph. The

node from which all branches start is called the root node. The other nodes are called subsidiaries.

One of the most promising solutions, in our opinion, are ontologies (Parveen, 2018). For example, we can use hierarchies with multi-criteria techniques during the classification of metadata of various articles. Ontologies aim to capture the domain knowledge generally and ensure a common understanding of the domain.

IsaViz is a virtual environment for viewing and creating RDF models in the view graphs. IsaViz imports RDF/XML and N-Triples and exports RDF/XML. Apollo is the program for modelling knowledge systems. Apollo's knowledge system base consists of hierarchically organized ontologies that can be inherited from other ontologies. SWOOP contains OWL (Web Ontology Language) validation and offers various. OWL presentation syntax views. In SWOOP, Ontologies can be compared, edited, and combined. Protégé 3.5 is a knowledge-based ontology editor that provides a graphical user interface. It ensures better flexibility for meta-modelling and enables the construction of domain ontologies.

1.4 Ontological Problems

Nowadays, most standard systems (such as Mendeley, Scopus etc.) provide support for displaying data but not for comparison and providing search functions. Also, given that articles in the same domain have the same indicators, the metadata of the results can be represented as ontology node attributes and then processed.

Previously, ontological graphs were used to systematize scientific articles (Parveen, 2018; Amami et al., 2017; Boughareb et al., 2020; Perraudin, 2017; Poulakakis et al., 2017). Systematization and structuring in such ontological systems were based on different approaches, such as using of scientific article recommendation system (Amami et al., 2017), a Scientific Articles Tagging system (Boughareb et al., 2020), machine learning and automatic summarization (Parveen, 2018). However, none of the proposed ontological approaches (Parveen, 2018; Amami et al., 2017; Boughareb et al., 2020; Perraudin, 2017) can provide a decent level of structurization and systematization.

We have proposed to use the cognitive IT platform Polyhedron (Stryzhak et al., 2019; Velychko et al., 2017; Strizhak, 2014) for this aim. The core of the Polyhedron system consists of advanced and improved functions of the TODOS IT platform described in previous works. Moreover, the Polyhedron is a multi-agent system that provides transdisciplinary

and interactivity in any study (Stryzhak et al., 2014). Besides, the cognitive IT platform Polyhedron contains a different variety of special functions like auditing (Stryzhak et al., 2014; Globa et al., 2015, 2019), semantic web, information systematization and ranking (Nadutenko et al., 2022), transdisciplinary support (Nicolescu, 2008; Dovgyi and Stryzhak, 2021), internal search and have all advantages of ontological interface tools (Globa et al., 2015; Popova and Stryzhak, 2013; Martyniuk et al., 2021) Due to active states are hyper-ratio plural partial ordering (Nicolescu, 2008; Volckmann, 2007), cognitive IT-platform Polyhedron is an innovative IT technology of ontological management of knowledge and information resources, regardless of the standards of their creation.

The proposed solution can be used with other applications in the field of structuration studies like a virtual educational experiment (Slipukhina et al., 2019), the use of mobile Internet devices (Modlo et al., 2019), using the technology of augmented reality education (Bilyk et al., 2022; Nechypurenko et al., 2023), and smart physiological tools (Shapovalov et al., 2022), distance learning in vocational education and training institutions (Kovalchuk et al., 2023), on-line courses (Vlasenko et al., 2020; Vakaliuk et al., 2023), educational and scientific environments (Tarasenko et al., 2021b; Shapovalov and Shapovalov, 2021), different tools to provide development of ICT (Information computer technologies) (Modlo et al., 2018).

2 MATERIALS AND METHODS

2.1 Ontology Creation Mechanism

Systematization of the scientific studies is provided using the cognitive IT platform Polyhedron (Shapovalov et al., 2021b). Graphs were made in one of two ways: using a single Google sheet file (that defines both structure of the graph and metadata of each node; or by using of specific constructor https://editor.stemua.science/ or http://work.inhost.com.ua/ that provided a generation of the XML file of both, structure and metadata. The first Google sheet document contained the data on the graph's structure. The second sheet of it contained a list of nodes that contains metadata, metadata of this node itself and the type of metadata (figure 2). When the file was filled, it was downloaded in .xlsx format and uploaded to the graph editors. Then, it was transformed into an XML file and stored in the graph's storage. This method was used to create massive graphs as it simplifies the creation of the graph structure due to no need

to select the location of the nodes visually with the mouse but locates nodes automatically. Tools https://editor.stemua.science/ or http://work.inhost.com.ua/ are graphical editors used to create simple graphs and test if the structure and metadata of graphs built by Google sheet's method are processed well.

As the scientific works were written with an IM-RAD structure, IMRAD was used as the core of each graph. For example, the field of anaerobic treatment of waste was chosen as general studies in this field also used IMRAD (Shapovalov et al., 2021a; Zhadan et al., 2021; Ivanov et al., 2019). To test using the approach for both high educational institutions students and schools' pupils (in the form of work of Junior Academy of Sciences of Ukraine), examples of such studies were used. As the example of master's study "Development a method for utilization of anaerobic digestion effluent at LLC Vasylkivska Poultry Farm" (study "A"); as an example of school student's study "Development a method for utilization of anaerobic digestion effluent (study "B"). Both studies were represented in the form of graphs.

To represent ontology graphs, the structure view in the form tools of view https://manlab.stemua.science/ and http://work.inhost.com.ua/ were used. In addition, ontology representer https://editor.stemua.science/ or http://work.inhost.com.ua/ has a structure view, table view and specific ontology prism view.

Metadata of generated graphs were used in filtering, ranking and audit tools. The simplest toll is filtering. It provides filtering of graph nodes by the presence of metadata in the nodes. The audit tool provides comparing of scientific studies with a standard graph. For science, this tool may provide two aims: define if such element of study (object, subject, keywords, etc.) was used previously to do not duplicate existing study; to provide simple (basic) antiplagiarism.

The ranking tool was used to compare numeric data in nodes and provide rank using users' requests. IT Polyhedron has a specific interface to input the importance of each metadata category that's numeric using a scale from 0 to 10. System process user's request by using values inputted by users and numeric data of nodes. Then it formed the rank of each node and showed it to the user.

To combine graphs, the first graph of scientific study that is IMRAD-based was downloaded to http://work.inhost.com.ua/ and then its tool "add XML" was used to add nodes and the link between them of the second IMRAD-based ontology graph of scientific study.

Figure 2: Google sheet with data.

3 RESULTS

3.1 Model of Creation Ontology to Systemizing

As was noted before, IMRAD is widely used to prepare research and science papers. It is possible to provide structuration by using IMRAD components as parent nodes. So, the parent nodes can be represented by Introduction, Methods, Results and Discussion. An ontology cannot efficiently further structure the discussion part. Important part that it contains is the analysis and the comparison of the obtained data by the researcher. Specific parts of IMRAD will be used as branch nodes, and the study will be used as a leaf node. So, the general structure of the ontology that structures the research data is represented as:

$$REP \in I, M, R, P \qquad (1)$$

where REP – set of reports, I – sets of Introduction of all study, M – set of Methods of all study, R – set of Results of all study, P – instruments of processing of the results of a set of studies discussions.

To provide better systematization, we have split the introduction into two different parts – basic metadata and literature review:

$$I = \langle BMD, LR \rangle \qquad (2)$$

where BMD – is a set of basic metadata of study, LR – a set of Sources used for Literature Review.

The primary metadata node of the study is linked with the graph's leaf nodes that characterize primary data on the study, such as hypothesis, object, subject, practical value, scientific novelty, etc. So, the nodes of the report's basic metadata of the study can be presented as a further equation:

$$BMD = \langle H, O, S, PV, SN \rangle \qquad (3)$$

where H – hypothesis or hypotheses of each specific study; O – object of the study of each specific study; S – the subject of each specific study; PV – practical value of each specific study; SC – the scientific novelty of each specific study.

The main advantages of using such a structure are that some parts of the Introduction materials and methods and results (measured parameters) of the study (reports) can coincide. A few specific studies that coincide will be linked by nodes (in case of methods and results) or by classes of data (in case of keywords or scientific novelty) due to using the same sub-nodes of the ontology. The representation of each work as a set of the Introduction, Methods, Results, and Processing of the data (Discussion):

$$REP_I = \langle I_I, M_I, R_I, P_I \rangle \qquad (4)$$

$$REP_{II} = \langle I_{II}, M_{II}, R_{II}, P_{II} \rangle \qquad (5)$$

So, these studies can be integrated into a single ontology using IMRAD:

$$O = \langle S_I, S_{II} \rangle = \langle I_I, M_I, R_I, P_I, I_I, M_I, R_{II}, P_{II} \rangle \qquad (6)$$

The same approach will be applied to each element of the IMRAD structure study. Generally, it can be represented as:

$$M = (REP_I) = \sum_i^n M_I \qquad (7)$$

where M_I – every separated scientific method.

In a different study, a different set of methods can be used. However, some of them can coincide. Thus, the set of methods used in two different studies may be represented as:

$$M = (REP_I) = M_a, M_b, M_c, M_d \qquad (8)$$

$$M = (REP_I I) = M_b, M_d, M_f \qquad (9)$$

And, so, M_b coinciding with both M_I and M_{II}:

$$M_b \in M_I, M_{II} \qquad (10)$$

Therefore, in this case, and M_b can be used as a parent node that connects two different studies. The node M_b itself will contain general theoretic information on it, and node REP_I and REP_{II} will contain information on the specific case of its usage and measured parameters using it.

Similar mechanism can be provided by using specific ontology tools using metadata. For example, there will be a hierarchical approach for representing and usage of keywords:

$$Kw(BMD_i) = Kw_a, Kw_b, Kw_c, Kw_d, \qquad (11)$$

where $Kw(BMD_i)$ – node of the basic metadata that integrates all keywords; Kw_i – specific keyword.

Also, as was noted in the introduction, the metadata of each work will be used for filtering the information, and for supporting specific processing functions of the IT solution Polyhedron. Such specifics mechanisms are AUDIT and RANKING. Metadata can be included in each node. For the parent node metadata will be used to represent the general information (for example, the essence of the method itself), and the resulting leaf node will contain the specific metadata related to a specific study (such as specific results of the study obtained using set methods M; for example, metadata: 5,35, and its class: "Ammonium nitrogen content, g/L"). So, metadata, with the same class, will be processed by using filtering by users request or by ranking using the ranks of the nodes for specific classes (or their set) based on the user's request.

So, the proposed approach uses IMRAD to collect and process the data with ontologies. In this way, the ontologies are constructed not by the specific structure of each work but by the generally accepted IMRAD structure. The parent node will be a specific area set to which the study belongs ($A = \sum inREP_{II}$ where A – specific area of the set of REP). The A node is linked with I, M, R, P nodes (representing IMRAD). Each IMRAD node is linked with the specific IMRAD type node (such as ammonia determination by Nessler's method (for methods) or "chicken manure" or "glycerine" (for subjects)). Moreover each specific IMRAD type node is linked with leaf nodes of ontology – specific studies where such entities were used.

3.2 Structuring of the Set of Studies in the Form of Ontology

To demonstrate the capabilities of the proposed ontological system, scientific works on anaerobic digestion were chosen. The general view of the resulting graph is shown in figure 3.

The root node of the resulting graph is the "Scientific reports" node. The ontological graph is complex because it has additional branches from child nodes in its structure. Child nodes are: "Reports on biotechnology" and "Reports on anaerobic digestion." From the child node, "Reports on anaerobic digestion" are going to the central sub-leaf nodes that reflect the basic principle of systematization of scientific works: "Results," "Materials and methods," and "References". This basic principle is shown in figure 4. A separate node of Main Metadata was also additionally created. This node contains the central metadata: object, subject of study; practical significance, the scientific novelty of study; hypotheses; keywords; abstract, conclusions.

The entire sequence and principle of filling and maintenance of data by users in the received ontology are shown in the Workflow diagram (figure 5).

The child nodes of these systematizing ontological nodes are the scientific works themselves. Each data block is in the form of separate attributes of the ontological node. This solution allows using all the information processing tools of the CIT Polyhedron system. In particular, such tools are general (for example, filtering) and specialized, such as ranking and auditing. An example of filtering is shown in figure 6.

An ontology prism is a specific form of ontology graph in cognitive IT Polyhedron. It provides the possibility to use nodes and their relation to the form of a prism. The most helpful form can represent the relevel of IMRAD and provide a high level of visualization of its sub-nodes. An example of an ontology cube representation of a graph is shown in figure 7.

3.3 Application of Ranking Mechanism in the Structuring of Scientific Works

All attributes can rank information using the "Alternative module" described in previous works (Nadutenko et al., 2022; Tarasenko et al., 2021a). The attributes of Each node are filled with numeric, textual, and mixed types of data. The following attributes are filled with text data: "References", "Methods for Quantitative Analysis", "Materials for researching", "Thermophilic", "Chicken manure substrate", "Spec-
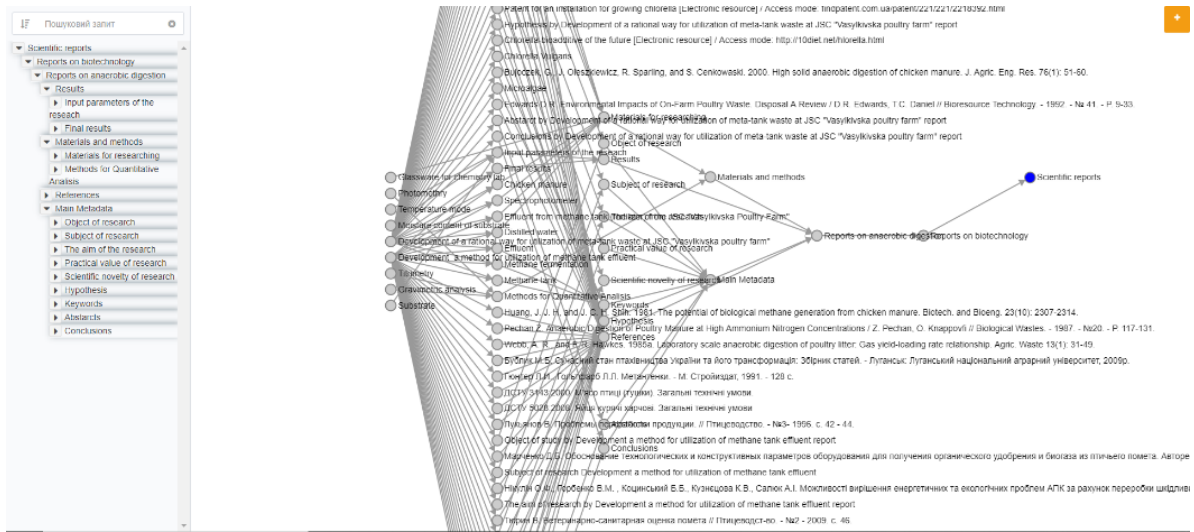
Figure 3: General view of the resulting ontological graph.

Figure 4: General view of the primary systematizing ontological vertices.

trophotometer parameters of the experiment", "The actual rate of reproduction of ", "Keywords", "Glassware", "Reagents", "Equipment", "Object ", "Subject of study", "The aim of the study", "Chicken". Numeric data contains the following attributes: "Initial pH", "Methane content, % Vol.", "pH of obtained solid product", "Ammonium nitrogen concentration, mg/L", "The concentration of volatile fatty acids (VFA) mg/L", "The dry matter content, %", "The ash content, %". The attribute "The native moisture content of the substrate" contains mixed-type data, numeric and textual. An example of incoming data maintenance panel for ranking is shown in figure 8.

For example, there may be a case when the user wants to arrange work on the pH. The ranking result is shown in figure 9. Other examples of usage of the Polyhedron IT platform are shown in table 1.

3.4 Application of the Audit Mechanism in Structuring Scientific Works

Users can also use the specialized audit module described in previous works (Stryzhak et al., 2014; Globa et al., 2015, 2019; Tarasenko et al., 2021a) for all attributes. The graph "Standard" is the ontology itself, containing works that will be supplemented and expanded. This solution will allow users to automatically check whether there is a particular work in the database. Also, this solution will allow checking the hypotheses for compliance with already completed studies. Also, the audit module will allow users to compare an existing METADATA and attributes available in the ontology at the same time. In particular, these attributes are the materials and methods of the results and the list of sources. Results that do not match the "standard" ontology attributes are em-
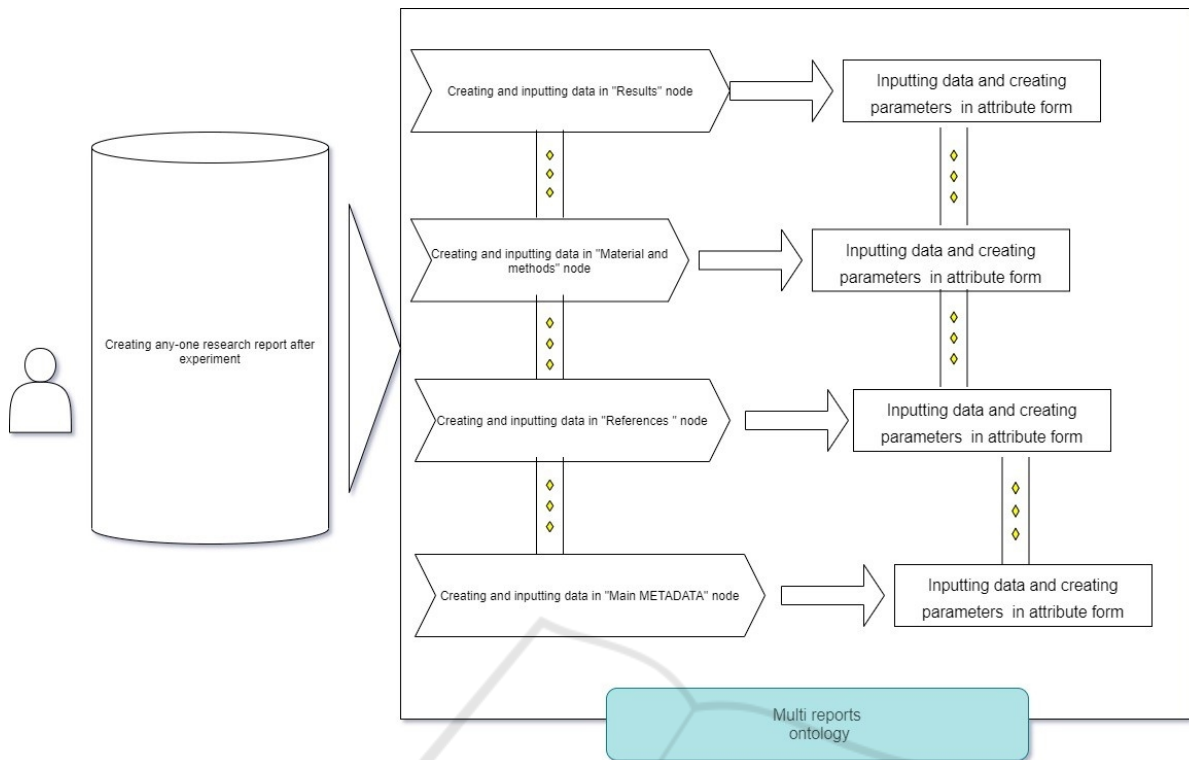
Figure 5: Workflow diagram of filling and maintenance of data by users.
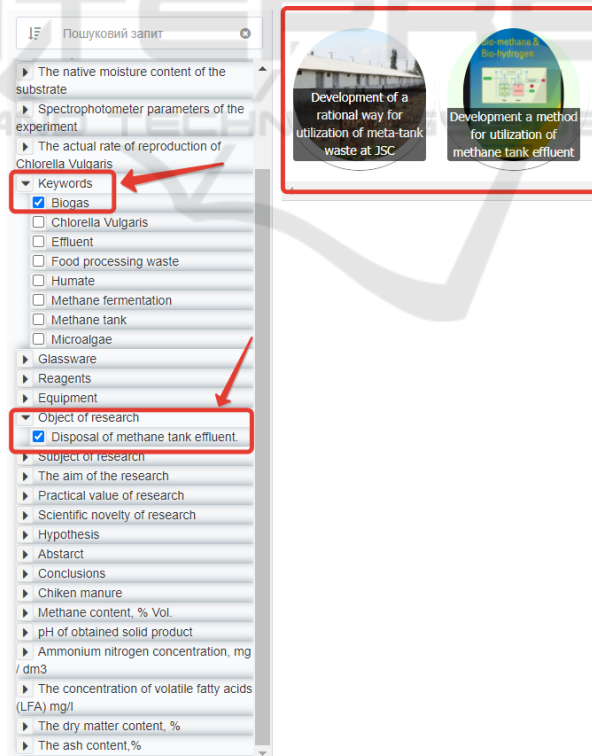

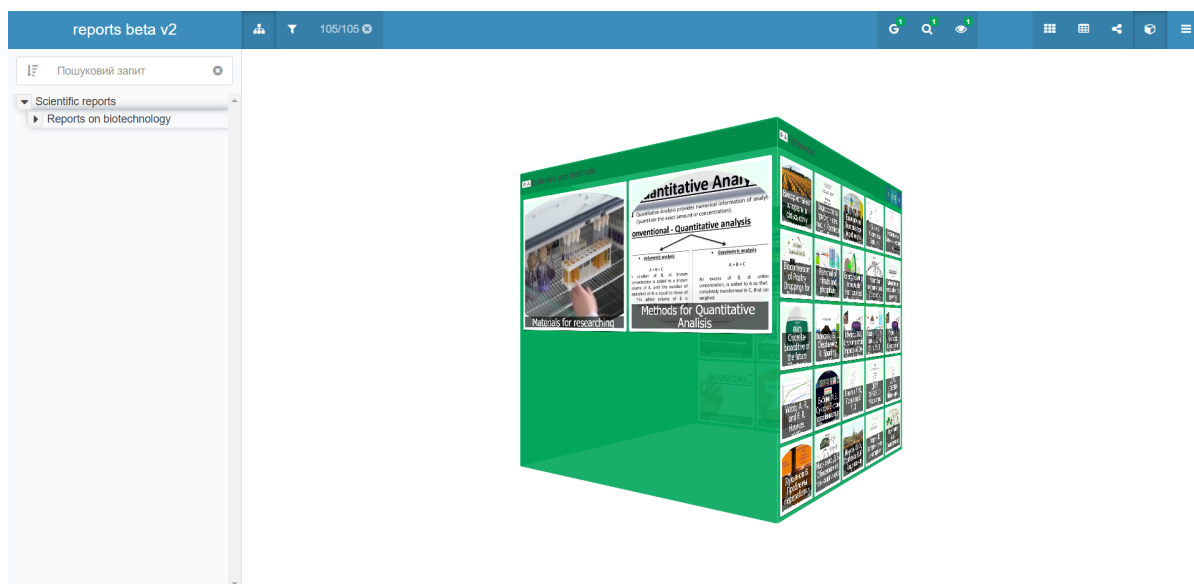Figure 6: Example of attribute filter.

Figure 7: Ontology prism view of IMRAD-based graph of scientific studies.

Figure 8: An example of incoming data maintenance panel for ranking.

phasized in red. An example of an audit fragment is shown in figure 10.

3.5 Interoperability of Ontology Graphs on Students' Scientific Studies

To make an effective personalized system, it is essential to provide interoperability between similar content of previous studies to allow users to create unique resources that will be used as the base of knowledge or literature review of the current researcher. So, interoperability is required. Moreover, using an IMRAD structure allows merging graphs because of the similar parent's nodes.

So, each student can use the developments made earlier. In addition, students can download a few existing ontology graphs from open data located on ontology.stemua.science and provide merging using MS Excel tool or specifically developed tools.

To solve this task, graphs-blanks (standards) are relevant to use. The standard consists of broad used categories for the field that will be filled by researchers who conduct studies. The general view of such graph-standard is presented in figure 11.

The researchers could input the specific data used during the research. Different researchers would have their specific graphs, but they were built by the same structure using IMRAD standard. Such graphs can be used to combine a single graph. Thus, the interoperability of different studies is provided. Therefore,
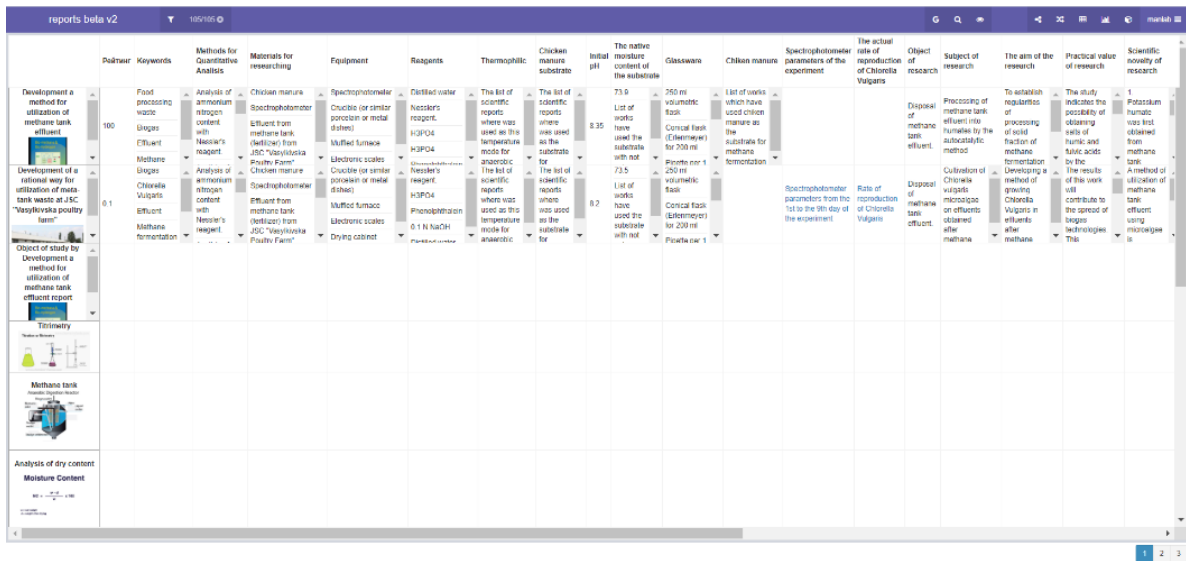
Figure 9: Example of ranking results.
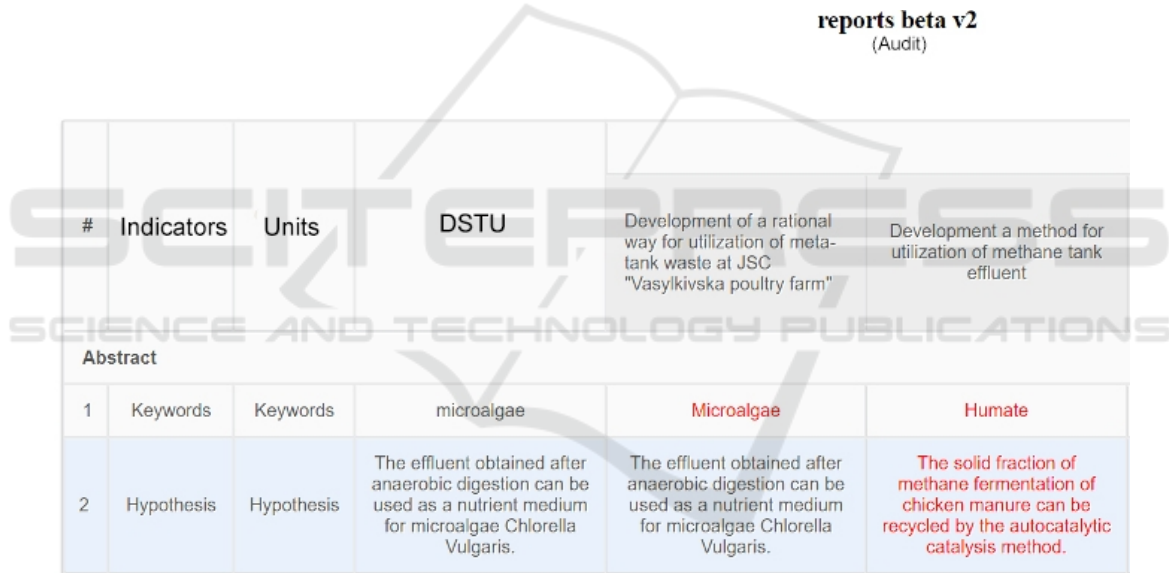
Figure 10: Example of audit results.

such graphs could contain more data on specific fields in the future.

As described in Section 3.1, the graph's combining will provide stability of IMRAD structure and the addition of sub nodes belonging to a specific study and the second to second. The two examples of filled studies to merge are presented in figure 12.

For example, the study "Anaerobic treatment under low moisture content" has a node "Biogas" related to node "Keywords". The same node also belongs to the second study, "Recirculation of the liquid phase of effluent during anaerobic digestion". After combining, the edge "Biogas" will have two subn-

odes "Anaerobic treatment under low moisture content" and "Recirculation of the liquid phase of effluent during anaerobic digestion" which provides belonging the keyword "Biogas" to both studies.

As for some other nodes that do not belong to both studies, they will also be represented in a combined graph, but they will contain only a single study to which it belongs. For example, such a case will be for nodes "Inhibition", "Ammonia", "Recycling", and "Water consumption". A general view of the result of combining is presented in figure 13.

(a)   (b)

Figure 11: The structure of the ghaph, (a) The function of adding to the graph another graph (combining), (b).

Table 1: Examples of Polyhedron IT Platform Ranking Module.

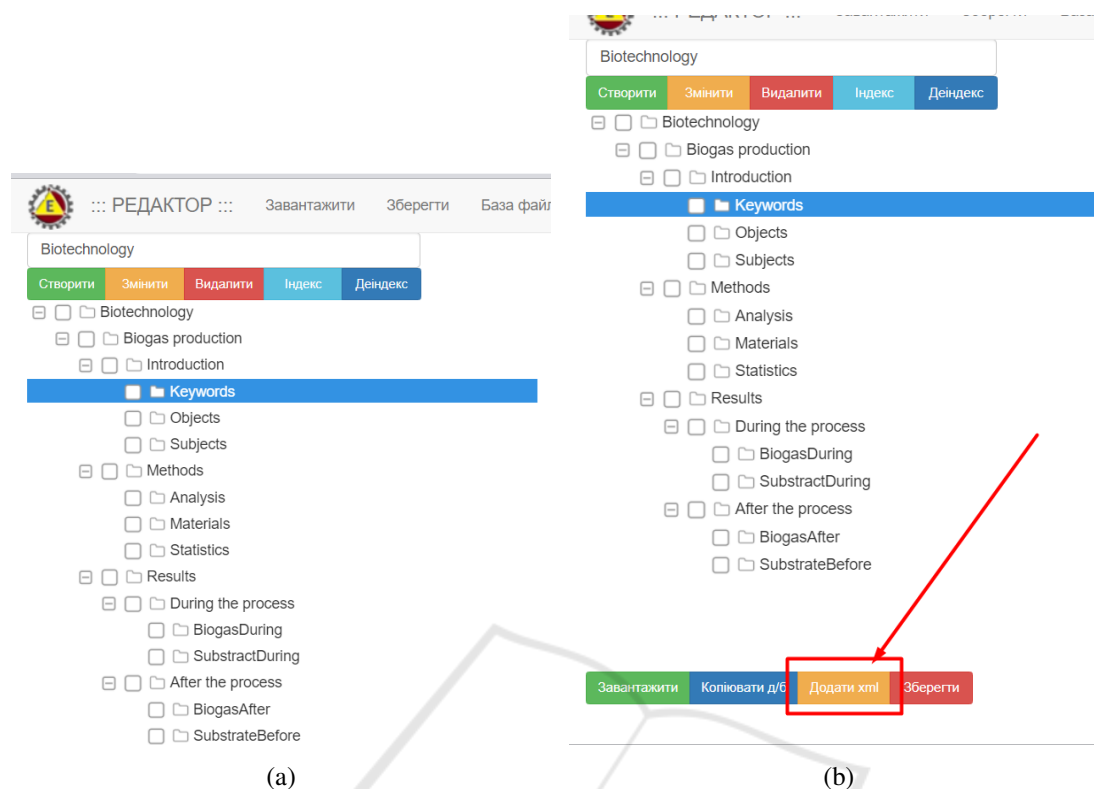| Name of filter (Ontological attributes)	Priority	Main results (list of ontological nodes)
Initial pH (type of data are numbers)	Absolute	Development of a rational way for utilization of meta-tank waste at JSC "Vasylkivska poultry farm", Titrimetry, Methane tank, Development a method for utilization of methane tank effluent Methods for Quantitative Analysis, Materials for researching, Abstract
Methane content, % Vol. (type of data are numbers)	Absolute	Development of a method for utilization of methane tank effluent, Development of a rational way for utilization of meta-tank waste at JSC "Vasylkivska poultry farm, Methods for Quantitative Analisis, Materials for researching, Abstract
The ash content, %	Absolute	Development of a rational way for utilization of meta-tank waste at JSC "Vasylkivska poultry farm, Development of a method for utilization of methane tank effluent, Methods for Quantitative Analisis, Materials for researching, Equipment, Reagents, Thermophilic, Chicken manure substrate
The dry matter content, %	Absolute	Development of a method for utilization of methane tank effluent, Development of a rational way for utilization of meta-tank waste at JSC "Vasylkivska poultry farm Keywords, Methods for Quantitative, Analysis, Materials for researching, Equipment, Thermophilic Chicken manure substrate

4 DISCUSSION AND CONCLUSIONS

4.1 Cases of Usage

Generally, it is possible to use such an approach in two cases "Personalized science study graph cre-

ation" and "Combining using Script editor". The first case foresees a simple scientific process, but the results of such a study and its description are provided by the creation ontology of such work using the graph template based on IMRAD. So, all researchers will prepare the description of their work using similar for-
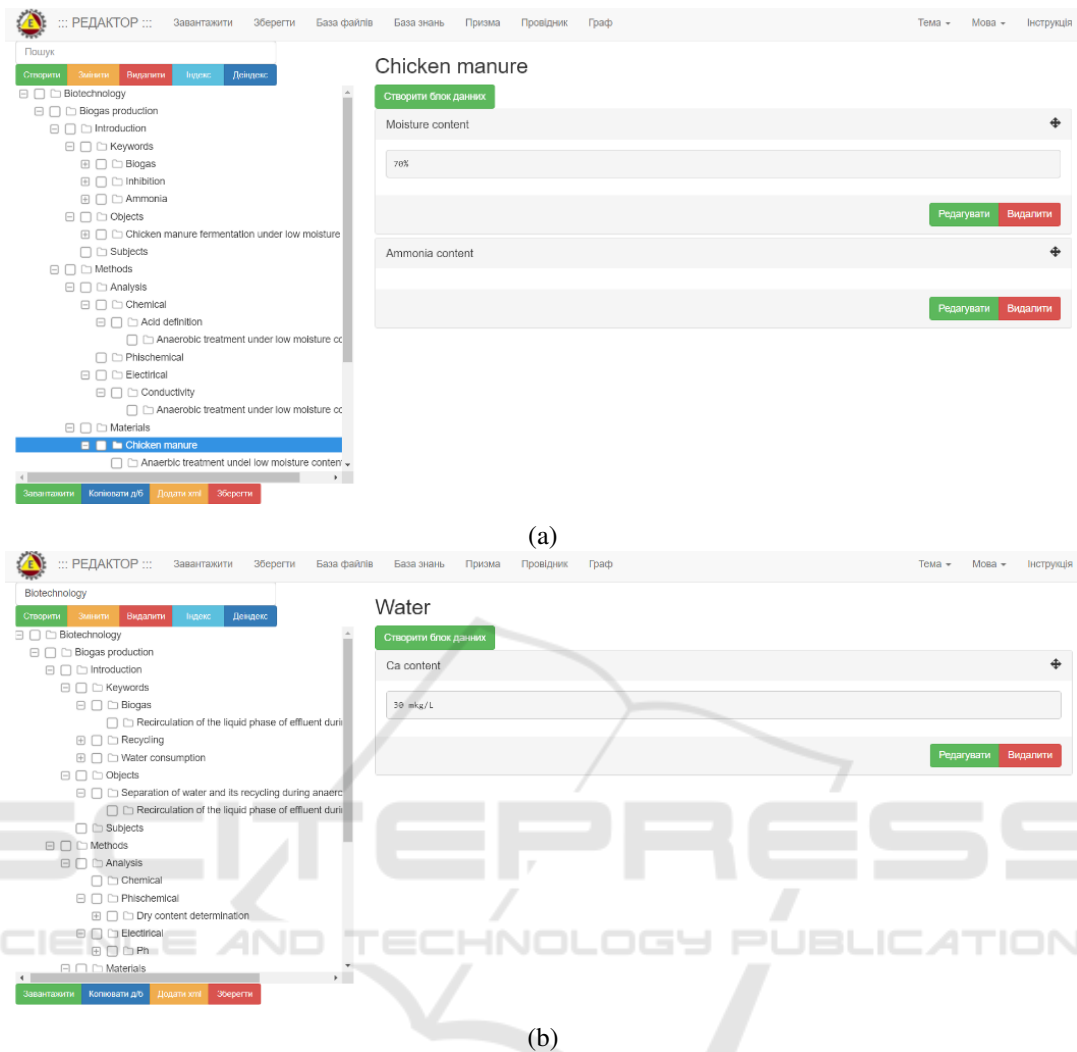
(a)

(b)

Figure 12: Example of ontology "Anaerobic treatment under low moisture content", (a) Example of ontology "Recirculation of the liquid phase of effluent during anaerobic digestion" (b).

malization. Furthermore, it provides the possibility to implement the "Combining using Script editor" use case. This use case is an easy way for the researcher to systemize all studies in his field that were done before. After combining, the obtained graph is merged as it was created from graphs with the same formalization. This graph is ready to use systemized literature review of the study for the scientist who provided it. Tools of Polyhedron will provide data processing of merged graph by using ranking, filtering and audit.

4.2 Role of Proposed System Among Existing Systems

Using of IMRAD as the primary approach to structure the articles in the form of a semantic ontology is proposed. The implementation of a universal ontological solution that can provide systematization and structuration of any scientific studies as proof of the concept is provided. The advantages and potential scenarios for using our solution have been demonstrated by the example of biogas studies. The potential of using ranking and auditing tools in the obtained ontological database has also been shown. Numeric and semantic characteristics were separated from the main text and used to process by specialized algorithms of IT Platform Polyhedron. For example, users can find studies where a specific method was used by both, using the structure and filtering of studies data. The numeric data of studies are processed by the ranking tool that assigns ranks to nodes depending on the value of these numeric characteristics. The created ontology allows to obtain the structured set of studies, separate their characteristics, provides the possibility to view
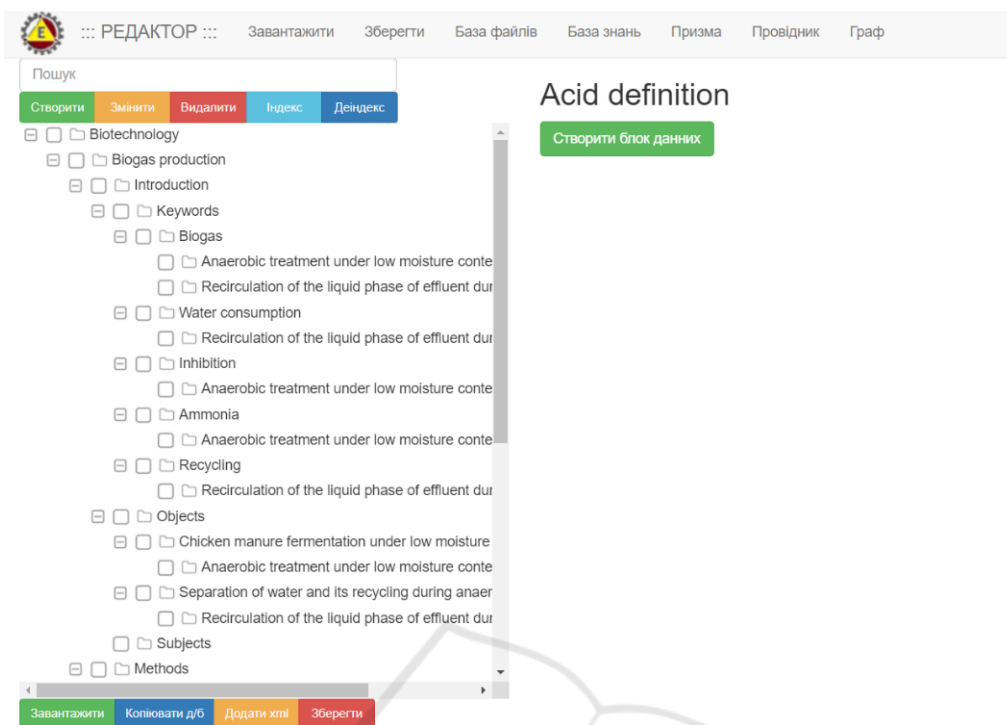
Figure 13: Combined graph of two studies.
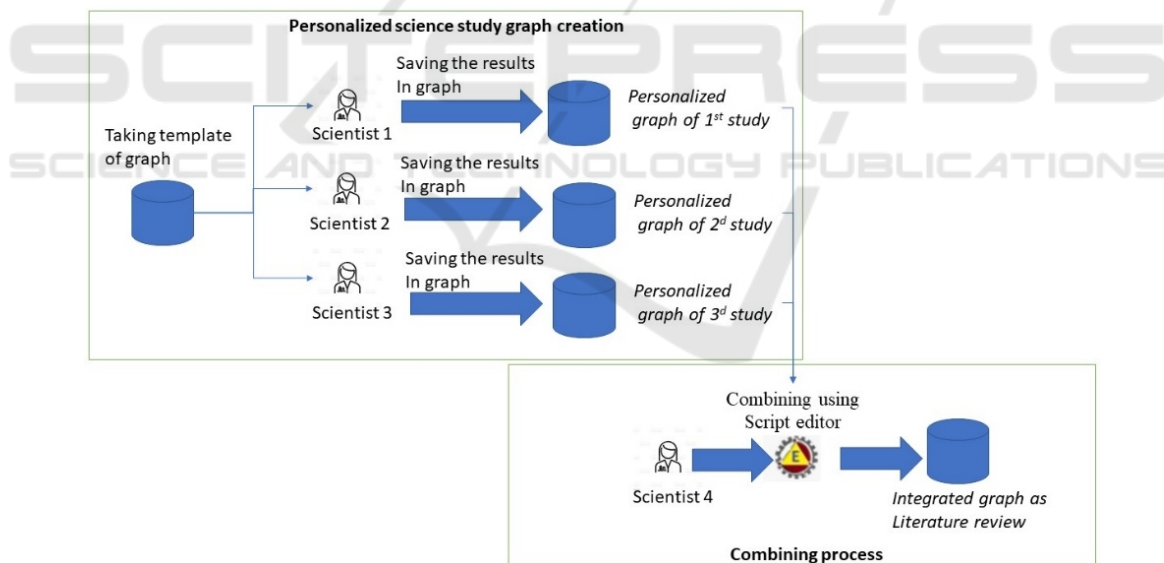
Figure 14: Use cases of the proposed approach.

all of the methods, measured parameters in the view of node and use them to find the studies where they were used. A detailed comparison of our ontological solution with the most common analogues is presented below.

As we can see from the table 2, proposed ontological solution has the essential functions of the most common software. Proposed solution can be used as bibliographic software and a scientometric database. In addition, our solution could provide such functions as ranking based on specific attributes, comparing different articles, visualizing the information given an ontological tree or ontology-based cube, and using the IMRAD approach to sort articles.

Mathematical interpretation of proposed model is based on that IMRAD foresees those papers consist

Table 2: Comparison of "Polyhedron" system with similar analogues.

	Personalized science study graph creation	Combining graphs using Script editor	Mendeley	Endnote	Scopus	Google Scholar
Automatic extraction of the information from any added PDFs	Present	Present	Present	Present	Absent	Absent
Tags, keywords, or search the full text of most PDFs functions.	Present	Present	Present	Present	Present	Present
Ability to cite articles in word/pages	Absent	Absent	Present	Present	Absent	Absent
Count citing	Possible	Present	Absent	Absent	Present	Present
Ability to use numeric data of the articles for ranking	Present	Present	Absent	Absent	Absent	Absent
Accentuation of important semantic characteristics for management of the wide range of articles	Present	Present	Absent	Absent	Absent	Absent
Ability to compare different articles	Possible	Present	Absent	Absent	Absent	Absent
Visualization of the information	Present	Present	Absent	Absent	Absent	Absent
Usage of IMRAD approach to sort articles	Present	Present	Absent	Absent	Absent	Absent

*proposed approach using CIT " Polyhedron"

for, Introduction, Methods, Results and Discussion. In form of detail systems, it is impossible to provide Discussion, but the similar results will be provided by data Processing. In turn, the formed graph of IMRAD that formed subject area is linked with specific to each study parts of each element of IMRAD (specific method or result). Therefore, the graph that describes few studies will consist from set of specific to each study parts of each element of IMRAD.

Such same formalization for different studies provides possibility to use specific features of ontologies. One of the most important is interoperability that is possible by merging few graphs of studies that has similar structure into single one. This feature provides possibilities to make digital ontology-based libraries of studies.

4.3 Further Research

The proposed approach may be much developed. The proposed approach foresees using Google Sheets and then transferring it into XML. However, it seems possible to simply such a process by generating ontologies exetly from Google Sheets and providing real-life synchronization.

Also, as the proposed approach is based on specific studies, it seems relevant and possible to provide scientometry. CIT "Polyhedron" is not a reference manager like "Endnote", "Mendeley", and it is also not a scientometric database, but in the nearest perspective, it is possible to convert the proposed technology into a functional analogue.

REFERENCES

Amami, M., Faiz, R., Stella, F., and Pasi, G. (2017). A Graph Based Approach to Scientific Paper Recommendation. In *Proceedings of the International Conference on Web Intelligence*, WI '17, page 777–782, New York, NY, USA. Association for Computing Machinery. https://doi.org/10.1145/3106426.3106479.

Bilyk, Z. I., Shapovalov, Y. B., Shapovalov, V. B., Megalinska, A. P., Zhadan, S. O., Andruszkiewicz, F., Dołhańczuk-Śródka, A., and Antonenko, P. D. (2022). Comparing Google Lens Recognition Accuracy with Other Plant Recognition Apps. In Semerikov, S., Osadchyi, V., and Kuzminska, O., editors, *Proceedings of the 1st Symposium on Advances in Educational Technology - Volume 2: AET*, pages 20–33. INSTICC, SciTePress. https://doi.org/10.5220/0010928000003364.

Boughareb, D., Khobizi, A., Boughareb, R., Farah, N., and Seridi, H. (2020). A Graph-Based Tag Recommendation for Just Abstracted Scientific Articles Tagging. *International Journal of Cooperative Information Systems*, 29(03):2050004. https://doi.org/10.1142/S0218843020500045.

Butros, A. and Taylor, S. (2011). Managing information: evaluating and selecting citation management software, a look at EndNote, RefWorks, Mendeley and Zotero. In *Proceedings of the 36th IAMSLIC Conference: Mar del Plata, Argentina, 17-21 October, 2010*, pages 53–66. https://hdl.handle.net/1912/4595.

Cuschieri, S., Grech, V., and Calleja, N. (2019). WASP (Write a Scientific Paper): The use of bibliographic management software. *Early Human Development*, 128:118–119. https://doi.org/10.1016/j.earlhumdev.2018.09.012.

Dovgyi, S. and Stryzhak, O. (2021). Transdisciplinary Fundamentals of Information-Analytical Activity. In Ilchenko, M., Uryvsky, L., and Globa, L., editors, *Advances in Information and Communication Technology and Systems*, volume 152 of *Lecture Notes in Networks and Systems*, pages 99–126, Cham. Springer International Publishing. https://doi.org/10.1007/978-3-030-58359-0_7.

Globa, L., Kovalskyi, M., and Stryzhak, O. (2015). Increasing Web Services Discovery Relevancy in the Multi-ontological Environment. In Wiliński, A., Fray, I. E., and Pejaś, J., editors, *Soft Computing in Computer and Information Science*, volume 342 of *Advances in Intelligent Systems and Computing*, pages 335–344, Cham. Springer International Publishing. https://doi.org/10.1007/978-3-319-15147-2_28.

Globa, L. S., Sulima, S., Skulysh, M. A., Dovgyi, S., and Stryzhak, O. (2019). Architecture and Operation Algorithms of Mobile Core Network with Virtualization. In Ortiz, J. H., editor, *Mobile Computing*, chapter 6. IntechOpen, Rijeka. https://doi.org/10.5772/intechopen.89608.

Ivanov, V., Stabnikov, V., Stabnikova, O., Salyuk, A., Shapovalov, E., Ahmed, Z., and Tay, J. H. (2019). Iron-containing clay and hematite iron ore in slurry-phase anaerobic digestion of chicken manure. *AIMS Materials Science*, 6(5):821–832. https://doi.org/10.3934/matersci.2019.5.821.

Ivey, C. and Crum, J. (2018). Choosing the Right Citation Management Tool: Endnote, Mendeley, Refworks, or Zotero. *Journal of the Medical Library Association*, 106(3):399–403. https://doi.org/10.5195/jmla.2018.468.

Kovalchuk, V., Maslich, S., and Movchan, L. (2023). Digitalization of vocational education under crisis conditions. *Educational Technology Quarterly*. https://doi.org/10.55056/etq.49.

Kumar Basak, S. (2014). A Comparison of Researcher's Reference Management Software: Refworks, Mendeley, and EndNote. *Journal of Economics and Behavioral Studies*, 6(7):561–568. https://doi.org/10.22610/jebs.v6i7.517.

Martyniuk, D., Falkenthal, M., Karam, N., Paschke, A., and Wild, K. (2021). An Analysis of Ontological Entities to represent Knowledge on Quantum Computing Algorithms and Implementations. In Paschke, A., Rehm, G., Qundus, J. A., Neudecker, C., and Pintscher, L., editors, *Proceedings of the Conference on Digital Curation Technologies (Qurator 2021), Berlin, Germany, February 8th - to - 12th, 2021*, volume 2836 of *CEUR Workshop Proceedings*. CEUR-WS.org. https://ceur-ws.org/Vol-2836/qurator2021_paper_15.pdf.

Modlo, Y. O., Semerikov, S. O., Nechypurenko, P. P., Bondarevskyi, S. L., Bondarevska, O. M., and Tolmachev, S. T. (2019). The use of mobile Internet devices in the formation of ICT component of bachelors in electromechanics competency in modeling of technical objects. *CTE Workshop Proceedings*, 6:413–428. https://doi.org/10.55056/cte.402.

Modlo, Y. O., Semerikov, S. O., and Shmeltzer, E. O. (2018). Modernization of Professional Training of Electromechanics Bachelors: ICT-based Competence Approach. In Kiv, A. E. and Soloviev, V. N., editors, *Proceedings of the 1st International Workshop on Augmented Reality in Education, Kryvyi Rih, Ukraine, October 2, 2018*, volume 2257 of *CEUR Workshop Proceedings*, pages 148–172. CEUR-WS.org. https://ceur-ws.org/Vol-2257/paper15.pdf.

Nadutenko, M., Prykhodniuk, V., Shyrokov, V., and Stryzhak, O. (2022). Ontology-driven lexicographic systems. In Arai, K., editor, *Advances in Information and Communication*, volume 438 of *Lecture Notes in Networks and Systems*, pages 204–215, Cham. Springer International Publishing. https://doi.org/10.1007/978-3-030-98012-2_16.

Nechypurenko, P., Semerikov, S., and Pokhliestova, O. (2023). Cloud technologies of augmented reality as a means of supporting educational and research activities in chemistry for 11th grade students. *Educational Technology Quarterly*. https://doi.org/10.55056/etq.44.

Nicolescu, B., editor (2008). *Transdisciplinarity: Theory and Practice*. Hampton Press, 1 edition.

Noyer, U., Beckmann, D., and Köster, F. (2009). Semantic technologies and metadata systematisation for evaluating time series in the context of driving experiments. In *11th International Protégé Conference*, pages 17–18. https://protege.stanford.edu/conference/2009/abstracts/P9-Noyer.pdf.

Parabhoi, L., Seth, A. K., and Pathy, S. K. (2017). Citation Management Software Tools: a Comparison with Special Reference to Zotero and Mendeley. *Journal of Advances in Library and Information Science*, 6(3):288–293. http://jalis.in/pdf/6-3/Arabinda.pdf.

Parveen, D. (2018). *A Graph-based Approach for the Summarization of Scientific Articles*. PhD thesis, Ruprecht-Karls-Universität Heidelberg. https://doi.org/10.11588/heidok.00027924.

Perraudin, N. (2017). *Graph-based structures in data science: fundamental limits and applications to machine learning*. PhD thesis, École Polytechnique Fédérale De Lausanne. https://infoscience.epfl.ch/record/227982?ln=en.

Popova, M. and Stryzhak, O. (2013). Ontological interface as a means of presenting information resources in the GIS. *Uchenye zapiski Tavricheskogo natcionalnogo universiteta imeni V. I. Vernadskogo. Seriia "Geografiia"*, 26(65)(1):127–135. https://tinyurl.com/5fhtrmck.

Poulakakis, Y., Vassilakis, K., Kalogiannakis, M., and Panagiotakis, S. (2017). Ontological modeling of educational resources: a proposed implementation for greek schools. *Education and Information Technologies*, 22(4):1737–1755. https://doi.org/10.1007/s10639-016-9511-z.

Ram, S. and Paul Anbu K., J. (2014). The use of bibliographic management software by Indian library and information science professionals. *Reference Services Review*, 42(3):499–513. https://doi.org/10.1108/RSR-08-2013-0041.

Salem, J. and Fehrmann, P. (2013). Bibliographic Management Software: A Focus Group Study of the Prefer-

ences and Practices of Undergraduate Students. *Public Services Quarterly*, 9(2):110–120. https://doi.org/10.1080/15228959.2013.785878.

Shapovalov, Y., Tarasenko, R., Usenko, S. A., Shapovalov, V., Andruszkiewicz, F., and Dołhańczuk-Śródka, A. (2021a). Ontological information system for the selection of technologies for the treatment and disposal of organic waste: engineering and educational aspects. *Desalination and Water Treatment*, 236:226–239. https://doi.org/10.5004/dwt.2021.27689.

Shapovalov, Y. B., Bilyk, Z. I., Usenko, S. A., Shapovalov, V. B., Postova, K. H., Zhadan, S. O., and Antonenko, P. D. (2022). Using Personal Smart Tools in STEM Education. In Semerikov, S., Osadchyi, V., and Kuzminska, O., editors, *Proceedings of the 1st Symposium on Advances in Educational Technology - Volume 2: AET*, pages 192–207. INSTICC, SciTePress. https://doi.org/10.5220/0010929900003364.

Shapovalov, Y. B. and Shapovalov, V. B. (2021). A Taxonomic Representation of Scientific Studies. In Ermolayev, V., Esteban, D., Mayr, H. C., Nikitchenko, M., Bogomolov, S., Zholtkevych, G., Yakovyna, V., and Spivakovsky, A., editors, *Proceedings of the 17th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I: Main Conference, PhD Symposium, and Posters, Kherson, Ukraine, September 28 - October 2, 2021*, volume 3013 of *CEUR Workshop Proceedings*, pages 353–360. CEUR-WS.org. https://ceur-ws.org/Vol-3013/20210353.pdf.

Shapovalov, Y. B., Shapovalov, V. B., Tarasenko, R. A., Usenko, S. A., and Paschke, A. (2021b). A semantic structuring of educational research using ontologies. *CTE Workshop Proceedings*, 8:105–123. https://doi.org/10.55056/cte.219.

Slipukhina, I., Kuzmenkov, S., Kurilenko, N., Mieniailov, S., and Sundenko, H. (2019). Virtual Educational Physics Experiment as a Means of Formation of the Scientific Worldview of the Pupils. In Ermolayev, V., Mallet, F., Yakovyna, V., Mayr, H. C., and Spivakovsky, A., editors, *Proceedings of the 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I: Main Conference, Kherson, Ukraine, June 12-15, 2019*, volume 2387 of *CEUR Workshop Proceedings*, pages 318–333. CEUR-WS.org.

Strizhak, A. E. (2014). Ontologicheskie aspekty transdistciplinarnoi integratcii informatcionnykh resursov [Ontological aspects of transdisciplinary integration of information resources]. *Otkrytye informatcionnye i kompiuternye integrirovannye tekhnologii*, 65:211–223. http://nbuv.gov.ua/UJRN/vikt_2014_65_24.

Stryzhak, O., Prychodniuk, V., and Podlipaiev, V. (2019). Model of Transdisciplinary Representation of GEOspatial Information. In Ilchenko, M., Uryvsky, L., and Globa, L., editors, *Advances in Information and Communication Technologies*, volume 560 of *Lecture Notes in Electrical Engineering*, pages 34–

75, Cham. Springer International Publishing. https://doi.org/10.1007/978-3-030-16770-7_3.

Stryzhak, O. Y., Horborukov, V. V., Franchuk, O. V., and Popova, M. A. (2014). Ontology selection problem and its application to the analysis of limnological systems. *Ecological safety and nature management*, 15:172–183. http://nbuv.gov.ua/UJRN/ebpk_2014_15_21.

Tarasenko, R. A., Shapovalov, V. B., Usenko, S. A., Shapovalov, Y. B., Savchenko, I. M., Pashchenko, Y. Y., and Paschke, A. (2021a). Comparison of ontology with non-ontology tools for educational research. *CTE Workshop Proceedings*, 8:82–104. https://doi.org/10.55056/cte.208.

Tarasenko, R. A., Usenko, S. A., Shapovalov, Y. B., Shapovalov, V. B., Paschke, A., and Savchenko, I. M. (2021b). Ontology-based learning environment model of scientific studies. In Kiv, A. E., Semerikov, S. O., Soloviev, V. N., and Striuk, A. M., editors, *Proceedings of the 9th Illia O. Teplytskyi Workshop on Computer Simulation in Education (CoSinE 2021) colocated with 17th International Conference on ICT in Education, Research, and Industrial Applications: Integration, Harmonization, and Knowledge Transfer (ICTERI 2021), Kherson, Ukraine, October 1, 2021*, volume 3083 of *CEUR Workshop Proceedings*, pages 43–58. CEUR-WS.org. https://ceur-ws.org/Vol-3083/paper278.pdf.

Vakaliuk, T., Chyzhmotria, O., Chyzhmotria, O., Didkivska, S., and Kontsedailo, V. (2023). The use of massive open online courses in teaching the fundamentals of programming to software engineers. *Educational Technology Quarterly*. https://doi.org/10.55056/etq.37.

Velychko, V., Popova, M., Prykhodniuk, V., and Stryzhak, O. (2017). TODOS – IT-platform formation transdisciplinaryn information environment. *Systems of Arms and Military Equipment*, (1(49)):10–19. http://nbuv.gov.ua/UJRN/soivt_2017_1_4.

Vlasenko, K. V., Volkov, S. V., Kovalenko, D. A., Sitak, I. V., Chumak, O. O., and Kostikov, A. A. (2020). Web-based online course training higher school mathematics teachers. *CTE Workshop Proceedings*, 7:648–661. https://doi.org/10.55056/cte.420.

Volckmann, R. (2007). Transdisciplinarity: Basarab Nicolescu Talks with Russ Volckmann. *Integral Review*, (4):73–90. https://tinyurl.com/39y6fdm3.

Zhadan, S., Shapovalov, Y., Tarasenko, R., and Salyuk, A. (2021). Development Of An Ammonia Production Method For Carbon-Free Energy Generation. *Eastern-European Journal of Enterprise Technologies*, 5(8-113):66 – 75. https://doi.org/10.15587/1729-4061.2021.243068.