

# Mathematical Interpretation of Educational Student's and Scientific Studies in Form of Digital Ontologies

Viktor B. Shapovalov<sup>a</sup> and Yevhenii B. Shapovalov<sup>b</sup>

*The National Center "Junior Academy of Sciences of Ukraine", 38-44 Degtyarivska Str., Kyiv, 04119, Ukraine*

**Keywords:** Ontology, IMRAD, Structuration, Scientific Studies, Biogas.

**Abstract:** Because of the problem of the large amount of scientific data generated, it is relevant to develop structuring and processing methods. Using ontology graphs is the modern perspective way of their representation. Considering that most studies are written based on IMRAD, it was used to provide integration of different studies at a single structure and provide structuration at all. The different ways to create integrated ontology using IMRAD are described. To get the necessary level of abstraction, IMRAD elements as part of a set of the specific studies were decomposed as levels of abstraction from L1 (general node that describes a branch of science) to L5 (specific papers with detailed data) depending on the abstraction. The content of each node in the form of metadata and its further processing is described. The particular way of using proposed modes has been described in the example of describing studies in biogas production. The mathematical model of the proposed ontology is developed and presented. It is shown that a set of corteges describes IMRAD representation of the scientific studies in the form of ontologies.

## 1 INTRODUCTION

The data nowadays is generated with colossal intensity. Due to this, Big Data processing is a trend (Globa et al., 2019; Stryzhak et al., 2021). Processing a considerable amount of data in real life is complicated by the high gain of publishing scientific studies. In general, it seems like an exponentially growing of the publications. According to lens.org, in 1900, only 532 M of scientific papers were published, but their amount in 2015 was near 10 B (figure 1).

Considering the development of STEM, studies are provided not only by experienced scientists by youth. Such a considerable number of studies generated complicated tasks to process such data. One of the problems of low spreading and usage (in the example of Ukraine (Hrynevych et al., 2021; Martyniuk et al., 2021; Shapovalov et al., 2020b; Stryzhak et al., 2017)) may be related to difficulties with the processing of science.

Now, scientific studies are published in different forms of report, such as articles, conference proceedings, books, etc. However, its process is complicated due to studies are low-structured. Sure, they are all

built by a similar structure named IMRAD (Oriokot et al., 2011; Pardede, 2012). It envisages requirements for the paper to consist of some generalized Introduction, describing used Materials and Methods, naming the Results of the study and the Discussion by comparing with other scientific materials or providing use cases. However, it seems not enough. Here just some examples of problems due to it:

- it is hard to start the researcher carrier due to complicated process of understanding of the methods and equipment that need to be used in specific fields of study;
- it is hard for youth scientists to understand main parameters that have measured to provide study analysis;
- for expired scientists, it is hard to analyze and collect data of new studies.

These are only very few cases that are a problem due to high amount of data of scientific studies. However, these cases are makes relevant to develop new methods to provide better structuration and data processing of scientific studies.

Sure, there are few solutions for this problem that provides automated science data processing (Klampfl et al., 2014; Portenoy and West, 2020; Gorashy and Salim, 2014; Shakeel et al., 2018; Paschke and

<sup>a</sup> <https://orcid.org/0000-0001-6315-649X>

<sup>b</sup> <https://orcid.org/0000-0003-3732-9486>

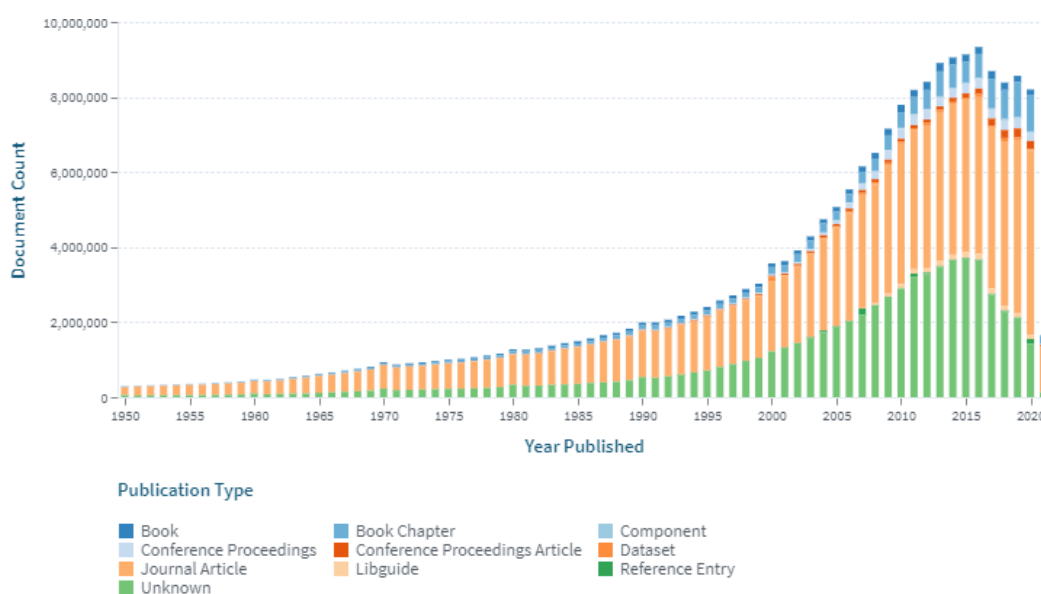


Figure 1: Dynamic of published papers according to Lens.org web service.

Schäfermeier, 2018), but it seems that they do not take to account IMRAD. One of the appropriate methods to solve the problems is ontology taxonomies (Globa et al., 2015; Mintser et al., 2018; Stryzhak et al., 2018; Schäfermeier et al., 2021) with semantic technologies (Alnemr et al., 2010). Also, ontology taxonomies have a lot of advantages, such as the possibility to combine with other types of materials (Gruber et al., 2020), including interactive and web-based courses (Bovtruk et al., 2020; Slipukhina et al., 2018), other information technologies (Markova et al., 2019; Modlo and Semerikov, 2018) and GIS (Stryzhak et al., 2019). This research aims to develop a model that can structure the set of the studies using IMRAD.

Previously, it was proposed to provide support using ontologies for single specific study, but not to create glossaries and structured sets of data. To provide it tools Open provenance, Ontology and EXPO (da Cruz et al., 2012) were developed. Another ontology solution in the field of science is MoKi that provides creation of wiki-based information scientific sources (Dragoni et al., 2014; Ghidini et al., 2012). There some specific ontology tools such as Gene ontology (Smith, 2008) or Centralized educational environment (Stryzhak et al., 2021). However, creation of ontology to structure the set of the studies seems relevant due lack of approaches to provide it.

## 2 METHODS OF THE RESEARCH

In the paper, the ontology model has developed using the main principles of Graph Theory, Set Theory,

and a Theory of Abstraction (Giunchiglia and Walsh, 1992). The graph was modelled using a simple hierarchical algorithm that foresees using only nodes and links. So, such a model further may be updated using the more comprehensive graph building tools such as weight coefficients. However, without simple modelling, providing it will not be possible. To provide structuration generally accepted structuring method IMRAD has been proposed and used.

To model data processing was developed taking to account the processing possibilities of the Polyhedron system due it has some advantages compare well known Protégé (The Board of Trustees of the Leland Stanford Junior University, 2020; Ameen et al., 2012) and OWL tools (Sinha and Couderc, 2012; Soldatova and King, 2006). Furthermore, the features of cognitive IT-platform tools Filtering, Audit, and Ranking to provide decision-making (Stryzhak et al., 2021; Shapovalov et al., 2019a,b) were described in equitations to describe the data processing in the ontology model.

## 3 RESULTS AND DISCUSSION

### 3.1 Using IMRAD to Provide Structure

As was noted before, IMRAD is used to prepare science papers. So, to provide structuration, it is possible to use parent nodes that represent IMRAD components. IMRAD – Introduction, Methods, Results, and Discussion. The discussion part can't be structured by

ontology because it contains the obtained data analysis and comparison. That is why discussion will be represented as the processing of the results.

$$(D \in S) \implies (P \in S) \quad (1)$$

where  $S$  – study (or set of studies),  $D$  – discussion of studies' results,  $P$  – processing of the results of a set of studies.

Approximate, ontology can be devoted to a specific field of science or integrate different fields. Depending on it, the ontology will have 5 or 4 abstract levels of deep. In the case of general ontology, the parent node will be “Scientific studies”, and its subsidiary nodes will name a specific field. In the case of a specific ontology, the parent node will name a specific field. Then it links with elements of IMRAD structure. Each element of IMRAD has its specific representation, and it's in turn linked with more specific for the study describing the element of IMRAD. And the leaf node will be a set of specific studies belonging to the field. Let's name each level with L symbols taking to account position in the hierarchy:

- L1 – General name of parent's node “Scientific studies”,
- L2 – Name of field of the study,
- L3 – Part of IMRAD,
- L4 – Specific representation of IMRAD (specific method, used materials, specific type of the results),
- L5 – Specific study where were used specific representations of IMRAD L4.

Therefore, the hierarchy in a specific study will have a form of {L2, L3, L4, L5} or the general ones will have a form of {L1, L2, L3, L4, L5}. Interoperability of the L2 nodes of two different graphs may be provided by using the graph constructor. It provides the possibility to merge graphs in two ways. The first foresees that graphs will be constructed as a general graph in the form of {L1, L2, L3, L4, L5} and with the same name of L1. And the second is to create L1 in the constructor and add there two specific graphs in the form of {L2, L3, L4, L5}. Schematic representation of the general ontology is shown in figure 2, and taxonomy of the specific field is shown in figure 3.

An alternative and a more humanly more human-readable way to provide abstraction are to revert this model and begin with L5 and end with L1. In this case, ontology will have structure form {L5, L4, L3, L2, L1}. The graph based on the abstraction that begins from specific studies L1 and ends by field of the research is shown in figure 4.

However, the main disadvantage of such a graph is evident and is the consequences of the structure:

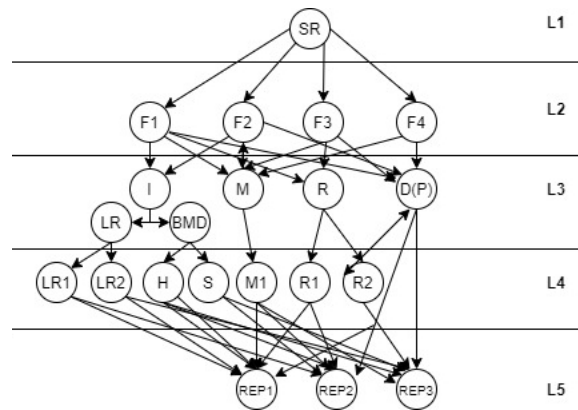


Figure 2: The taxonomy of the general science report ontology, where LR1, LR2, M1, R1, R2 – are abstract classes of literature review (LR), Methods and results of object.

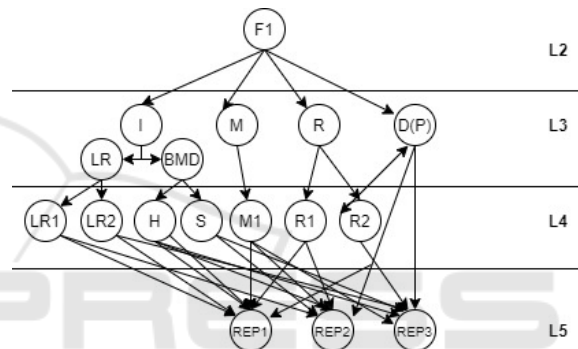


Figure 3: The taxonomy of the specific field science report ontology.

the leaf node SR (“Scientific study”) will be not very useful for users. Anyway, this type of graph may be built as {L5, L4, L3} and in this case, it will be used to evaluate the specific report, for example, during qualifying work evaluation (PhD or Master’s study). It will show abstract classes of each specific part of IMRAD for each specific study and can provide an evaluation of the set of methods and results that the researcher obtained. Anyway, in this research, we’ll use the first way to provide hierarchies in the form of {L2, L3, L4, L5}, and {L1, L2, L3, L4, L5}.

As it can be seen, the general science report ontology is significantly more complicated due to links between L1 and L2 levels, and also, there will be some problems with a vast amount of methods, results, etc. that can be not necessary to the user that looking for information on the specific field. Also, it will be much harder to create such type of graphs due it will have two levels of links “one to many” (see figure 2, links between L2 and L3 level and links between L4 and L5 levels) compare to only one in case of specific ontology (see figure 3, only links between L4 and L5 levels). It may be unreasonable to create a compli-

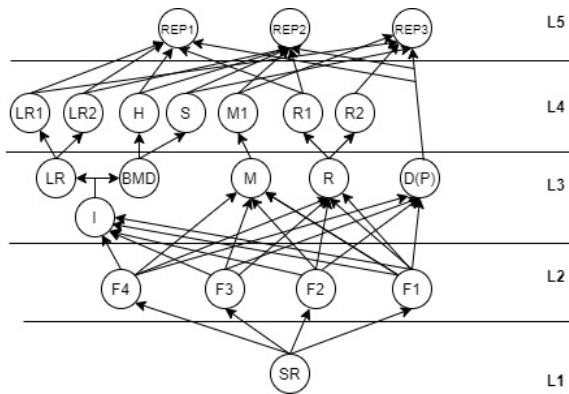


Figure 4: The graph based on the abstraction that begins from specific studies L1 and ends by field of the research.

cated graph. Therefore, it seems relevant to provide both types of hierarchies. To provide it, the ontologies should be created in specific fields and then merged, as noted before.

In this case, specific parts of IMRAD will be used as subsidiaries nodes in the field of the study, and specific studies will be used as leaf nodes. So, the general structure of such ontology may be represented as:

$$\{I, M, R, P\} \in REP \quad (2)$$

where  $I$  – sets of Introduction of all studies,  $M$  – set of Materials and Methods an of all studies,  $R$  – set of Results of all studies,  $P$  – processing of the results of a set of studies; replaces discussion;  $REP$  – report (or set of report).

To provide better systematization and we have split the introduction into two different parts due to their specific – basic metadata and literature review; it is possible to represent the introduction as further:

$$I = \langle BMD, LR \rangle \quad (3)$$

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

Basic metadata of the study node linked with graph nodes that characterized the essential data on the study, such as hypothesis, object, subject, practical value, and scientific novelty. And so, a node of the primary report's metadata of the study can be presented as a further equation:

$$BMD = \langle H_i, O_i, S_i, PV_i, S_i \rangle \quad (4)$$

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

Each work of the set of the Introductions, Methods, Results, and Processing of the data (Discussion). Then each work will be represented as the future:

$$S_I = \langle I_i, M_i, R_i, P_i \rangle \quad (5)$$

$$S_{II} = \langle I_{II}, M_{II}, R_{II}, P_{II} \rangle \quad (6)$$

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

$$\langle S_I, S_{II} \rangle = \langle I_i, M_i, R_i, P_i, I_{II}, M_{II}, R_{II}, P_{II} \rangle \quad (7)$$

### 3.2 Using Taxonomy Nodes as Structure of Science Data

The main advantages of using such structures are that some parts of the introduction (for example, keyword), materials and methods and results elements (entities and measured parameters) of studies/report in the same field can coincide and, in this case, such coinciding sub-nodes will be used as links for them and provide their interoperability. 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 to which a set of the studies belongs to ( $L2 = \sum_i^n RS_{I_i}$ , where  $L2$  – specific area and  $RS$  – set of the represented studies). The  $L2$  node is linked with  $I, M, R, P$  nodes (representing IMRAD). Each IMRAD node is linked with a specific node (such as ammonia determination by Nessler's method (for methods) or "chicken manure" or "glycerine" (for subjects)) that belongs to such types. And each specific IMRAD type is linked with leaf nodes of ontology – specific studies where such entities were used.

In this case, a few studies/report (REP1, REP2, and REP3 that belong to L5) will be integrated with some of the methods or results (M1, R1, R2 that belong to L4). So, the L4 level will be used to provide the structuration of the studies (L5). The user can use it in both ways: to find which method, result, etc., that belong to L4 were used in a specific report that belongs to L5; and define in which studies belong to L5 specific method, result, etc. that belong to L4 were used.

The same approach will be provided for each element of the structure. General can be represented as:

$$L4(M) = \sum_i^n M_i \quad (8)$$

where  $M_i$  – every separated scientific method.

Case of coinciding of the methods may be represented as single mortises of methods of each study:

$$M_I = \{M_a, M_b, M_c, M_d\} \quad (9)$$

$$M_{II} = \{M_b, M_d, M_f\} \quad (10)$$

Therefore, in this case,  $M_b$  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  $S_I$  and  $S_{II}$  will contain information on the specific case of its usage and measured parameters using it.

Also, for example, there will be a hierarchical way of representing and using the keywords:

$$K_w(BMD_i) = Kw_a Kw_b, Kw_c, Kw_d \quad (11)$$

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

In this case, some of the studies, same as for the methods,  $Kw_i$  will be elements of two different studies ( $Kw_a, Kw_c \in S_I, S_{II}$ ). This will be useful, especially for students and young scientists looking to find methods ( $M_I$ ) and parameters that can be used in specific fields and their usage in practice. Also, this way provides a list of the parameters and methods used in specific fields.

### 3.3 Metadata Processing

The metadata of each work will be used for processing the data. It may be included for each node. For example, metadata of L4 nodes will represent the general information (for example, the essence of the method itself), and the resulting leaf nodes 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 it’s class: “Ammonium nitrogen content, g/l). And so, metadata with the same class will be processed by filtering by users’ request or by ranking by providing the rank of nodes by specific class (or their set) based on the user’s request. So, each node located on each level  $E_i$  contains metadata with the abstract level that corresponds to several levels; for level 1st – it will be the most abstract metadata, and for 5th – it will be the most specific.

As can be seen, all data in levels L1-L4 contains generalized metadata and wouldn’t be used to process specific study, but just used to get generalized abstract information on entities used in specific fields. Only the L5 level contains metadata related to a specific study and will be used for further processing.

### 3.4 Using Metadata to Provide Data Processing

Specific mechanisms “Filtering”, “AUDIT” and “RANK” of cognitive IT solution Polyhedron are used to provide processing of the information. It will be used for the case when different studies will have the same Class and Type of information, but different

values:

$$\{Class : C1; Type : Number; Value : V1\} \in REP1 \quad (12)$$

$$\{Class : C1; Type : Number; Value : V2\} \in REP2 \quad (13)$$

$$\{Class : C1; Type : Number; Value : V3\} \in REP3 \quad (14)$$

And the values  $V_1, V_2, V_3$  can be equal or not equal. Anyway “Filtering”, “AUDIT” and “RANKING” can be used to process the data. Filtering can be described by function if:

If ( $V_{min} < V < V_{max}$ ) then (display nodes with such V)

or

If ( $V = V_{set}$ ) then (display nodes with such V) where  $V_{min}, V_{max}, V_{set}$  are maximum, minimum, and given (set) values, respectively, that inputted by the user.

The function of AUDIT can also be described as a function if:

If ( $V_i = V_{set}$ ) then (mark red such  $V_i$ ); for each  $V_i$ .

The ranking is much more complicated and can be described as:

$$RANK_{abs(i)} = \sum (OR_i \times IMP_i \times \frac{V_i}{V_{max}}) \quad (15)$$

where  $RANK_{abs(i)}$  – ranking rank in absolute value for  $i$ ’s node  $OR_i$  – orientation maximum or minimum for metadata of  $i$ ’s object (can be +1 or -1);  $IMP_i$  – importance coefficient for metadata of  $i$ ’s object;  $V_i$  – the value of metadata of  $i$ ’s object;  $V_{max}$  – maximum value of the set of metadata.

$$RANK_i = \frac{RANK_{abs(i)}}{RANK_{max}} \quad (16)$$

where  $RANK_i$  – the relative value of the rank (can be maximum =1) of each object;  $RANK_{max}$  – the maximum value of the RANK for all sets of objects.

### 3.5 Formalization Description

The object of formalization is specific scientific studies. The result of formalization is a specialized research-oriented subject area formed precisely from existing research and allows to familiarize with the specialized subject area. Any research essentially has the same components (which are proposed to be systematized in the form of graphs) – introduction (landscape, object of research, subject of research, novelty, etc.), methods (a set of methods that ensures the achievement of a scientific result or measurement), specific achievements and results (e.g., systems and approaches developed or metrics) and discussion. All components except the last one can be formalized using the IMRAD approach in such a way that they form an ontology of the subject area of a specific field of research. Discussion, in its essence, is finding the place

Table 1: Description of the metadata on each ontology of proposed ontology model.

<b>MD(L1)</b>	<b>no metadata</b>
<b>MD(L2)</b>	{Class: Information about the field; Type: String; Value: Description}
<b>MD(L3)</b>	[MD(LR); MD(BMS); MD(M); MD(R)] = LR, BMS, M, R {Class: General information; Type: String; Value: Describing and detailing of meaning results, methods, literature review, etc.}
<b>MD(L4)</b>	$\sum [MD(LR_i); MD(BMS_i); \{ MD(M_i); MD(R_i) \}] = \{ \text{Class: Essence of the name (specific method, results, etc.); Type: String; Value: Describing of way of providing or specific measured parameter} \}$
<b>MD(L5)</b>	$\sum \{ \text{Class: all metadata of specific study; Type: Number or String; Value: Text or number} \}$

of this research in the system of scientific research – that is, it is the process of comparing the results of research, numerical and other data with existing other data and providing explanations of the differences of this specific stud. In fact, such processing is provided by the ranking tools and the CIT Polyhedron alternative.

## 4 DISCUSSION

### 4.1 Case of Usage: An Example on Biogas Production

So, for the specific case of biogas production studies (Ivanov et al., 2019; Shapovalov et al., 2020a; Plyatsuk and Chernish, 2014; Bochmann et al., 2020), it seems relevant to use ontology for a specific field (in the form of {L2, L3, L4, L5}). In this case, a node in the L2 line will be single and named “Studies on anaerobic digestion”. It will be linked with nodes Introduction, Methods, Results, and Processing. As for all other cases, Introduction will be divided into Basic Metadata and Literature review (L3 level).

Basic Metadata will be linked with nodes Objects, Subjects, Aims, Practical Value, Scientific novelty, Hypothesis, Keywords, Abstract, Conclusion (L3 level).

Each of these nodes will be connected with specific nodes relevant to the set of the structured studies (L4 level). Each specific L3 will have metadata with general information on the described object. So, an example of values of metadata in the “Basic metadata” elements node in the L4 level is shown in table 2.

\*verbs “are defined” or “has provided” etc. and articles “the”, “a” and “an” aren't use due to their huge vitiation and to provide better structuration and to have more coincidences between nodes and metadata

Each such node will be connected with the study where it was used (L5 level). For example, “Biogas

production from the poultry waste” or “Utilization of the meat production wastewater using anaerobic digestion”.

The Literature review node (L4 level) will be connected with specific studies used in a set of studies. Its name will be the name of the study (paper, article, conference processing, thesis, etc.), similar to the name of the study used to provide structuration with the addition of the publishing year. For example, it can be named “Utilization of the meat production wastewater using anaerobic digestion, 2011”. In addition, each such node should be connected to one of the few studies used to provide structuration (L5 level).

The most useful will be Methods and Results nodes. They will be helpful to students and youth scientists who want to be familiar with methods used in the field and set the measured parameters used in the field of science. Sure, the established scholars will use such a tool too to increase outlook. The Materials and Methods node will be divided into Methods, Equipment, and Materials. An example of material and methods and results nodes, their links and metadata are presented in table 3.

Each such subsidiary node is connected with a leaf node that is a specific study. For example, the Processing node has metadata with type link and its value in the form of a link to Audit and Ranking tools for the structured set of studies. Detailed algorithms of its usage are described before.

Each work has metadata that mostly duplicates the structure. For this, all numeric and semantic data of the works is added to a node of the specific work it belongs to. Examples of the metadata of the leaf nodes are presented in the table. It is foreseen to provide automatically. For example, it will be necessary to provide filtering, Audit, and ranking. An example of metadata and its classes (subclasses) of the specific report node is shown in table 4.

Table 2: An example of “Basic metadata” elements nodes in L3 level and linked with them nodes in L4 level.

Parent’s node (L3)	Metadata of the parent’s node	Linked nodes (L4)
<b>Objects</b>	General definition of the elements of basic metadata	“biogas production”, “inhibition”, “waste utilization”
<b>Subjects</b>		“Effect of ammonium nitrogen content on biogas production”, “Optimization of the process of waste treatment by optimization of the waste destruction rate”
<b>Aims</b>		“Provide mathematical modeling of the anaerobic digestion of high-ammonium waste”, “Define of influence of the addition of spirulina to the process of anaerobic treatment of straw”
<b>Practical Value</b>		“Main kinetic parameters of the anaerobic digestion”, “Model of ammonia effect on the anaerobic digestion”
<b>Scientific novelty</b>		“Relation between ammonia content and biogas production”
<b>Hypothesis</b>		
<b>Keywords</b>		“Straw”, “Sludge”, “Meat wastewater”, “Biogas”, “Methane”, “Ammonium nitrogen”
<b>Abstract</b>		–
<b>Conclusion</b>		–

Table 3: An example of material and methods and results nodes links and metadata

Parent’s node	Metadata of the father’s node (type: text)	Linked subsidiary nodes	Metadata
Methods	General information what is methods	“Dry organic matter by frying”, “Methane content in biogas using gas chromatography”, “Free acid content by titrimetric method”	Methodology of using of each specific method (type: text)
Equipment	General information what is equipment	“Digital microscope”, “Burette”, “Gas chromatograph”	Description of each specific equipment (type: array) Link to ontology of the equipment (type: text)
Materials	General information what is material	“Straw”, “Sludge”, “Meat wastewater”, “Water”	Description of each specific materials (type: text)
Results	General information what is results	“Biogas”, “Methane”, “Ammonium nitrogen”	Description of each measured parameter (type: text)

## 4.2 Role of the Proposed Model

Ontology models are the basis of the effective ontology creative process. Such models like proposed and others (for example, ontologies of educational environments, will be useful to build a set of the different ontologies and have similar conceptual states of abstraction. Using such approaches and providing semantic technologies can be useful to provide interoperability (Alnemr et al., 2010).

Sure, the proposed research focused on the ontology of the specific field in the form of {L2, L3, L4, L5}, but it is proposed to use an integrator of the ontologies of fields and create general ontology in the form of {L1, L2, L3, L4, L5}. The proposed integration is important to provide transdisciplinary (Dovgyi and Stryzhak, 2021). The proposed approach

will be useful and relevant for most fields. Anyway, it will be very specific to process humanitarian data where less standardization and numeric data, but it seems that some automated tools like recursive reducer (Stryzhak et al., 2018) can process and provide structuration even in such fields.

## 4.3 Perspectives of Development

Currently, the proposed approach has a few user stories implemented by the proposed model. They are helpful for all scientists, but as the development of the proposed model was provided in the Navigational center Of Junior Academy of Sciences of Ukraine, it has much more advantages for youth students involved in activities of the organization. The mathematical interpretation of educational students and sci-

Table 4: An example of metadata and its classes (subclasses) of the specific report node.

Name of class	Name of subclass	Type	Values example
Methods	–	Array	“Dry organic matter by frying”, “Methane content in biogas using gas chromatography”, “Free acid content by titrimetric method”
Results	Biogas content, ml/ g TS	Number	“305.15”
	Methane content, %	Number	“55”
	Ammonium nitrogen content, g/l	Number	“3.6”
Materials	Straw/TS content, %	Number	“95”
	Straw/Ammonium nitrogen content, g/l	Number	“0.3”
	Sludge/TS content, %	Number	“0.05”
Main metadata	Keywords	Array	“Straw”, “Sludge”, “Meat wastewater”, “Biogas”, “Methane”, “Ammonium nitrogen”

entific studies in the form of digital ontologies provides the possibility to easily manage information of science studies to simplify finding of relevant studies and simplify familiarization process with some specific subject area.

The proposed approach

- 1) allows very quickly (especially for a young scientist) to research the subject field related to this field of research by using → Introduction → Keywords (contains the main terms of the subject field of specific research) and other components of the Introduction (for example, scientific novelty formulates the directions of research, which formulates relevant research directions);
- 2) allows to process numerical research data using the ranking tool and find such works that are necessary for research;
- 3) allows you to quickly familiarize with the existing research methods used in this field → Methods;
- 4) allows to quickly familiarize with the indicators used in research in a specific field (→ Results) Communication with L5 vertices is essential because it is he who forms the novelty (since the approaches to the ontological display of subject area have been known for a long time);
- 5) allows the researcher/student (young scientist) to quickly find practical examples where this or that element of research is used – for example, quickly find all works where ammonia was measured using the Nessler method or works where graph theory was used.

In addition, this approach has the potential for development, which is as follows:

- the possibility of providing scientometrics based on ontologies (similar to scientific databases) – since it is possible to calculate how many times

a particular work has been referred to due to the connections in such a taxonomy;

- the possibility of interoperability providing with educational programs;
- the possibility of adding one's own research for a few clicks to the general ontology.

## 5 CONCLUSIONS

It is firstly proposed the model of ontology based on IMRAD to provide a set of different studies that belong to the same field and to provide generation of the integrated ontology that collected the data of different fields. Using such a method will provide both structuration of the set of studies by using specific elements of IMRAD that belongs to the set of the studies of the same field and processing such studies' data.

A specific case of usage is shown in the example creation of such ontology in the field of biogas production. It is shown in both model and example using single sets of keywords, results, methods, etc., to provide structuring and data processing.

It seems relevant to provide additional further studies of the proposed model to improve it and make it even more automatized, for example, by using weight mechanisms.

The proposed approach in case of providing property infrastructure and widespread will provide interoperability of data located in papers. Therefore, it will simplify providing of scientific studies and simplify determination of relevance and practical value of scientific works. To provide such interoperability graphs of specific fields should be created and provided their further merging. So, the ontologies type {L2, L3, L4, L5} must be integrated into single one with form of {L1, L2, L3, L4, L5}.



## REFERENCES

- Alnemr, R., Paschke, A., and Meinel, C. (2010). Enabling Reputation Interoperability through Semantic Technologies. In *Proceedings of the 6th International Conference on Semantic Systems, I-SEMANTICS '10*, New York, NY, USA. Association for Computing Machinery. <https://doi.org/10.1145/1839707.1839723>.
- Ameen, A., Khan, K. U. R., and Rani, B. P. (2012). Creation of Ontology in Education Domain. In *2012 IEEE Fourth International Conference on Technology for Education*, pages 237–238. <https://doi.org/10.1109/T4E.2012.50>.
- Bochmann, G., Pesta, G., Rachbauer, L., and Gabauer, W. (2020). Anaerobic Digestion of Pretreated Industrial Residues and Their Energetic Process Integration. *Frontiers in Bioengineering and Biotechnology*, 8. <https://doi.org/10.3389/fbioe.2020.00487>.
- Bovtruk, A., Slipukhina, I., Mienailov, S., Chernega, P., and Kurylenko, N. (2020). Development of an electronic multimedia interactive textbook for physics study at technical universities. In Bollin, A., Mayr, H. C., Spivakovsky, A., Tkachuk, M. V., Yakovyna, V., Yerokhin, A., and Zholtkevych, G., editors, *Proceedings of the 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I: Main Conference, Kharkiv, Ukraine, October 06-10, 2020*, volume 2740 of *CEUR Workshop Proceedings*, pages 159–172. CEUR-WS.org. <https://ceur-ws.org/Vol-2740/20200159.pdf>.
- da Cruz, S. M. S., Campos, M. L. M., and Mattoso, M. (2012). A Foundational Ontology to Support Scientific Experiments. In Malucelli, A. and Bax, M. P., editors, *Proceedings of Joint V Seminar on Ontology Research in Brazil and VII International Workshop on Metamodels, Ontologies and Semantic Technologies, Recife, Brazil, September 19-21, 2012*, volume 938 of *CEUR Workshop Proceedings*, pages 144–155. CEUR-WS.org. [https://ceur-ws.org/Vol-938/ontobras-most2012\\_paper12.pdf](https://ceur-ws.org/Vol-938/ontobras-most2012_paper12.pdf).
- 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](https://doi.org/10.1007/978-3-030-58359-0_7).
- Dragoni, M., Bosca, A., Casu, M., and Rexha, A. (2014). Modeling, Managing, Exposing, and Linking Ontologies with a Wiki-based Tool. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1668–1675, Reykjavik, Iceland. European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2014/pdf/769\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/769_Paper.pdf).
- Ghidini, C., Rospocher, M., and Serafini, L. (2012). Conceptual Modeling in Wikis: a Reference Architecture and a Tool. In *Proceedings of the 4th International Conference on Information, Process, and Knowledge Management (eKNOW 2012)*, pages 128–135. [https://www.thinkmind.org/index.php?view=article&articleid=eknow\\_2012\\_6\\_10\\_60015](https://www.thinkmind.org/index.php?view=article&articleid=eknow_2012_6_10_60015).
- Giunchiglia, F. and Walsh, T. (1992). A theory of abstraction. *Artificial Intelligence*, 57(2):323–389. [https://doi.org/10.1016/0004-3702\(92\)90021-O](https://doi.org/10.1016/0004-3702(92)90021-O).
- 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 Pejaš, 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](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*, Rijeka. IntechOpen. <https://doi.org/10.5772/intechopen.89608>.
- Gorashy, Z. and Salim, N. (2014). Systematic literature review (SLR) automation: A systematic literature review. *Journal of Theoretical and Applied Information Technology*, 59(3):661–672.
- Gruber, M., Eichstädt, S., Neumann, J., and Paschke, A. (2020). Semantic Information in Sensor Networks: How to Combine Existing Ontologies, Vocabularies and Data Schemes to Fit a Metrology Use Case. In *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*, pages 469–473. <https://doi.org/10.1109/MetroInd4.0IoT48571.2020.9138282>.
- Hrynevych, L., Morze, N., Vember, V., and Boiko, M. (2021). Use of digital tools as a component of STEM education ecosystem. *Educational Technology Quarterly*, 2021(1):118–139. <https://doi.org/10.55056/etq.24>.
- 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/matiersci.2019.5.821>.
- Klampfl, S., Granitzer, M., Jack, K., and Kern, R. (2014). Unsupervised document structure analysis of digital scientific articles. *International Journal on Digital Libraries*, 14(3):83–99. <https://doi.org/10.1007/s00799-014-0115-1>.
- Markova, O. M., Semerikov, S. O., Striuk, A. M., Shalatska, H. M., Nechypurenko, P. P., and Tron, V. V. (2019). Implementation of cloud service models in training of future information technology specialists. *CTE Workshop Proceedings*, 6:499–515. <https://doi.org/10.55056/cte.409>.
- Martyniuk, O. O., Martyniuk, O. S., Pankevych, S., and Muzyka, I. (2021). Educational direction of STEM in the system of realization of blended teaching of physics. *Educational Technology Quarterly*, 2021(3):347–359. <https://doi.org/10.55056/etq.39>.
- Mintser, O. P., Pryhodnyuk, V. V., Stryzhak, O. Y., and Shevtsova, O. M. (2018). Transdisciplinary reporting of information with interactive documents. *Med-*

- ical Informatics and Engineering*, (1):47–52. <https://doi.org/10.11603/mie.1996-1960.2018.1.8891>.
- Modlo, Y. O. and Semerikov, S. O. (2018). Xcos on Web as a promising learning tool for Bachelor's of Electromechanics modeling of technical objects. *CTE Workshop Proceedings*, 5:34–41. <https://doi.org/10.55056/cte.133>.
- Oriokot, L., Buwembo, W., Munabi, I. G., and Kijjambu, S. C. (2011). The introduction, methods, results and discussion (IMRAD) structure: a Survey of its use in different authoring partnerships in a students' journal. *BMC Research Notes*, 4(1):250. <https://doi.org/10.1186/1756-0500-4-250>.
- Pardede, P. (2012). Scientific Articles Structure. In *Scientific Writing Workshop, The English Teaching Study Program of the Christian University of Indonesia (UKI), April 29–May 27, 2012*. <https://www.researchgate.net/publication/260453687>.
- Paschke, A. and Schäfermeier, R. (2018). OntoMaven - Maven-Based Ontology Development and Management of Distributed Ontology Repositories. In Nalepa, G. J. and Baumeister, J., editors, *Synergies Between Knowledge Engineering and Software Engineering*, pages 251–273, Cham. Springer International Publishing. [https://doi.org/10.1007/978-3-319-64161-4\\_12](https://doi.org/10.1007/978-3-319-64161-4_12).
- Plyatsuk, L. and Chernish, E. (2014). Intensification of Anaerobic Microbiological Degradation of Sewage Sludge and Gypsum Waste Under Bio-Sulfidogenic Conditions. *The Journal of Solid Waste Technology and Management*, 40(1):10–23. <https://doi.org/10.5276/JSWTM.2014.10>.
- Portenoy, J. and West, J. D. (2020). Constructing and evaluating automated literature review systems. *Scientometrics*, 125(3):3233–3251. <https://doi.org/10.1007/s11192-020-03490-w>.
- Schäfermeier, R., Herre, H., and Paschke, A. (2021). Ontology Design Patterns for Representing Context in Ontologies Using Aspect Orientation. In Blomqvist, E., Hahmann, T., Hammar, K., Hitzler, P., Hoekstra, R., Mutharaju, R., Poveda-Villalón, M., Shimizu, C., Skjæveland, M. G., Solanki, M., Svátek, V., and Zhou, L., editors, *Advances in Pattern-Based Ontology Engineering, extended versions of the papers published at the Workshop on Ontology Design and Patterns (WOP)*, volume 51 of *Studies on the Semantic Web*, pages 183–203. IOS Press. <https://doi.org/10.3233/SSW210014>.
- Shakeel, Y., Krüger, J., von Nostitz-Wallwitz, I., Lausberger, C., Durand, G. C., Saake, G., and Leich, T. (2018). (Automated) Literature Analysis: Threats and Experiences. In *Proceedings of the International Workshop on Software Engineering for Science*, SE4Science '18, page 20–27, New York, NY, USA. Association for Computing Machinery. <https://doi.org/10.1145/3194747.3194748>.
- Shapovalov, V. B., Shapovalov, Y. B., Bilyk, Z. I., Atamas, A. I., Tarasenko, R. A., and Tron, V. V. (2019a). Centralized information web-oriented educational environment of Ukraine. *CTE Workshop Proceedings*, 6:246–255. <https://doi.org/10.55056/cte.383>.
- Shapovalov, Y., Zhadan, S., Bochmann, G., Salyuk, A., and Nykyforov, V. (2020a). Dry anaerobic digestion of chicken manure: A review. *Applied Sciences*, 10(21):7825. <https://doi.org/10.3390/app10217825>.
- Shapovalov, Y. B., Shapovalov, V. B., Andruszkiewicz, F., and Volkova, N. P. (2020b). Analyzing of main trends of STEM education in Ukraine using stemua.science statistics. *CTE Workshop Proceedings*, 7:448–461. <https://doi.org/10.55056/cte.385>.
- Shapovalov, Y. B., Shapovalov, V. B., and Zaselskiy, V. I. (2019b). TODOS as digital science-support environment to provide STEM-education. *CTE Workshop Proceedings*, 6:235–245. <https://doi.org/10.55056/cte.382>.
- Sinha, A. and Couderc, P. (2012). Using OWL ontologies for selective waste sorting and recycling. In Klinov, P. and Horridge, M., editors, *Proceedings of OWL: Experiences and Directions Workshop 2012, Heraklion, Crete, Greece, May 27-28, 2012*, volume 849 of *CEUR Workshop Proceedings*. CEUR-WS.org. [https://ceur-ws.org/Vol-849/paper\\_16.pdf](https://ceur-ws.org/Vol-849/paper_16.pdf).
- Slipukhina, I. A., Olkhovyk, V. V., Kurchev, O. O., and Kapranov, V. D. (2018). Development of education and information portal of physics academic course: Web design features. *Information Technologies and Learning Tools*, 64(2):221–233. <https://doi.org/10.33407/itlt.v64i2.1781>.
- Smith, B. (2008). Ontology (Science). *Nature Precedings*. <https://doi.org/10.1038/npre.2008.2027.1>.
- Soldatova, L. N. and King, R. D. (2006). An ontology of scientific experiments. *Journal of The Royal Society Interface*, 3(11):795–803. <https://doi.org/10.1098/rsif.2006.0134>.
- Stryzhak, O., Horborukov, V., Prychodniuk, V., Franchuk, O., and Chepkov, R. (2021). Decision-making System Based on The Ontology of The Choice Problem. *Journal of Physics: Conference Series*, 1828(1):012007. <https://doi.org/10.1088/1742-6596/1828/1/012007>.
- 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](https://doi.org/10.1007/978-3-030-16770-7_3).
- Stryzhak, O. Y., Prychodniuk, V. V., Haiko, S. I., and Shapovalov, V. B. (2018). Vidobrazhennia merezhevoiv informatsii u vyhliadi interaktyvnykh dokumentiv. Transdystyplinarnyi pidkhyd. *Matematychni modeliuvannia v ekonomitsi*, (3):87–100. [http://nbuv.gov.ua/UJRN/mmve\\_2018\\_3\\_10](http://nbuv.gov.ua/UJRN/mmve_2018_3_10).
- Stryzhak, O. Y., Slipukhina, I. A., Polikhun, N. I., and Chernetckiy, I. S. (2017). STEM-education: Main definitions. *Information Technologies and Learning Tools*, 62(6):16–33. <https://doi.org/10.33407/itlt.v62i6.1753>.
- The Board of Trustees of the Leland Stanford Junior University (2020). protégé. <https://protege.stanford.edu/products.php>.