НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

UDC 004.9

LOGIC-ONTOLOGICAL RECONSTRUCTION OF SCIENTIFIC DISCOURSE AND ITS IMPLEMENTATION IN AN AI-BASED REVIEWING SYSTEM

Bedratyuk L. P. – Dr. Sc., Professor, Head of the Department of Software Engineering, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

ABSTRACT

Context. The growing number of scientific publications and the emergence of tools based on large language models (LLMs) highlight the need for automated verification of the structural quality of scientific texts. Most existing solutions focus on surface-level linguistic analysis and do not account for logical-discursive integrity – specifically, whether the text includes a hypothesis, method, results, conclusions, and whether these elements are connected by normative relationships.

Objective. The aim of this study is to develop an ontology-driven approach for the formalized verification of scientific text structures by constructing an ontological knowledge graph and evaluating its compliance with a predefined normative model of scientific discourse.

Method. A model is proposed based on two interrelated ontologies: “Scientific Publication” (defining node types and their roles) and “Reviewing” (defining logical-discursive requirements). The text is represented as a graph where nodes are formed through semantic markup using an LLM, and connections are verified according to a set of normative rules. A specialized GPT agent capable of dynamically applying ontological knowledge during analysis and review generation is employed for implementation.

Results. The model enables automatic detection of discourse structure violations: the absence of key elements, logical discontinuities, substitution of scientific novelty with practical significance, and incorrect interpretation of results. The proposed metrics quantitatively capture the level of structural completeness and consistency. Provided examples of graphs and reviews demonstrate that the system can detect non-obvious, latent logical inconsistencies even in formally complete texts.

Conclusions. The scientific novelty of the study lies in introducing the ontological graph as an interpretable model of scientific argumentation, used in tandem with a large language model. The practical significance lies in establishing a foundation for semi-automated reviewing, structural analysis of publications, and academic writing training. The methodology is scalable to other genres of scientific texts and can potentially be integrated into editorial platforms.

KEYWORDS: ontology, ontological graph, GPT agent, reviewing, semantic analysis, scientific publication, logical-discursive structure, machine learning, algorithms.

ABBREVIATIONS

BFO is a Basic Formal Ontology;

COPE is a Committee on Publication Ethics;

GPT is a Generative Pre-trained Transformer;

IMRaD is A standard structure of academic papers, dividing them into Introduction, Methods, Results, and Discussion sections;

LLM is a Large Language Model;

UFO is an Unified Foundational Ontology.

NOMENCLATURE

A is an artificially generated texts with controlled defects;

B is a publications from journals excluded from scientometric databases;

C is a benchmark articles;

E is a set of directed edges;

FN is a set of articles in which actual defects were missed;

FP is a set of articles in which non-defective elements were incorrectly flagged as defective;

G is a ontological knowledge graph;

L is a node labeling function;

s_i is a i-th fragment of a scientific text;

S is an ordered set of text fragments;

t_i is a i-th type of a node;

v_i is a i-th node in an ontological graph;

V is a set of nodes in an ontological graph;

V_{expected} is a set of ontological nodes expected in a reference annotation;

V_{correct} is a set of ontological nodes correctly identified by the system;

$\mathfrak{R}$ is a set of admissible relation types between nodes;

T is a set of ontological terms;

TN is a set of articles correctly identified as having no defects;

TP is a set of articles in which defects were correctly detected;

Φ is a set of logical-discursive rules;

DOI 10.15588/1607-3274-2025-4-7

$\phi_i(G)$ is a value of a Boolean predicate;

Δ is a set of violated rules.

INTRODUCTION

In the context of a growing volume of scientific publications and the widespread adoption of large language models (LLMs), the issue arises not only of ensuring linguistic correctness but also of maintaining the structural and logical coherence of scientific discourse. Peer review, as a key instrument of scholarly communication, requires significant time and cognitive resources, especially given the thematic diversity and the increasing number of manuscript submissions. This underscores the need for tools supporting preliminary or automated assessment based on a formal structural model of a scientific text.

Most existing solutions focus on superficial analysis or review generation, without considering the logical-discursive structure of the publication. They lack an explicit knowledge model of a scientific article as a system of interconnected semantic components – namely, hypothesis, method, results, and conclusions. Attempts to integrate ontologies into the analysis of scientific texts remain largely theoretical and do not cover the full pipeline – from semantic markup to the generation of a verifiable review. In contrast, modern LLMs enable the integration of generative capabilities with formal ontologies, paving the way toward interpretable and logically controlled reviewing systems.

The object of this study is the process of expert evaluation of scientific publications based on their logical-discursive structure.

The subject of the study is an ontological model of a scientific publication represented as a knowledge graph $G=(V,E,L)$ with fixed node types and normative links, as well as its integration into an automated reviewing process using a GPT-based agent.

The goal of this work is to design and implement an ontology-driven model for the analysis of scientific texts, which allows for the automatic generation of a preliminary review by detecting structural completeness, logical consistency, scientific novelty, argumentative soundness, and typical markers of pseudo-scientific imitation.

To achieve this goal, the following tasks must be accomplished:

– to construct an ontology of a scientific publication that reflects key nodes and the logical dependencies among them;

– to formulate a reviewing ontology as a set of predicates and normative relationships;

– to develop a review template aligned with both ontologies;

– to implement a GPT-based agent that performs semantic annotation, graph construction, and logical validation;

– to conduct an experiment with a controlled corpus and evaluate the accuracy of defect detection;

– to assess the applicability of the proposed approach in editorial and educational contexts.

The proposed approach combines the formal semantics of ontologies with the flexibility of generative language models, forming a new class of hybrid systems for structural-logical reviewing. The developed system not only identifies key structural elements of the text but also evaluates their coherence, scientific relevance, and compliance with the standards of academic reasoning. Such a tool can be useful to editors and reviewers of scientific journals for pre-screening submitted manuscripts, detecting potential structural flaws, and ensuring a more objective and rapid evaluation process. The implementation as a standalone GPT agent and the possibility of API integration open up perspectives for embedding the system into editorial platforms or using it in educational settings – particularly in courses on academic writing and scholarly communication.

The GPT agent is implemented within the ChatGPT environment and can assist journal editors and reviewers in the preliminary screening of submitted manuscripts, detection of potential structural deficiencies, and facilitation of more objective and efficient assessments.

1 PROBLEM STATEMENT

This study addresses the problem of logical-ontological analysis of a scientific article, represented as an ordered set of textual fragments $S=\{s_1,s_2,\ldots,s_n\}$, where each fragment s_i is a logically complete unit of scientific discourse. The ultimate objective is to construct a graph-based representation of the article's content in the form of an ontological graph $G=(V,E,L)$, where the set of nodes V corresponds to conceptual units of the text, the set of edges $E\subseteq V\times V$ reflects logical and argumentative relationships, and the labeling function $L:V\to T$ assigns each node an ontological type from a predefined set T (e.g., Hypothesis, Method, Result, Conclusion).

The input to the model includes the article text S, the set of ontological terms, and a set of verification rules $\Phi=\{\phi_1,\ldots,\phi_k\}$, each represented as a Boolean function applied to the graph G, enabling assessment of structural completeness and logical coherence.

The expected output of the task is an ontological graph G that reflects the logical-discursive structure of the scientific article. Additionally, the truth value $\phi_i\in\Phi$ is computed for each rule, and an aggregate consistency metric is calculated as the arithmetic mean of all fulfilled rules:

$$\text{Consistency}(G)=\frac{1}{k}\sum_{i=1}^{k}\phi_i(G).$$

The problem is formalized with consideration of several constraints. First, each textual fragment may be assigned no more than one ontological type from the set CCC.

Second, the graph must be connected or consist of connected components that include nodes of the key types – Hypothesis, Method, Result, and Conclusion. Third, the relations between nodes must conform to the

© Bedratyuk L. P., 2025
DOI 10.15588/1607-3274-2025-4-7

admissible links defined in the corresponding ontology, such as Method validates Hypothesis, Result supports Hypothesis, and Conclusion follows from Result.

2 REVIEW OF THE LITERATURE

The issue of knowledge formalization in the form of ontologies has become the subject of intensive research in philosophy, logic, and computer science. The fundamental principles of the ontological approach were formulated in works [1] and [2], which proposed upper-level ontologies (in particular, UFO) that formed the basis for semantic modeling in systems analysis. These works devote significant attention to the rigor of concept classification and the relationships between them. An extension toward natural language analysis is proposed in [3], where natural language ontology is considered as a means of formalizing the meaning of linguistic units within a philosophical-semantic context.

The issues of the truth of statements, truth-bearers, and the structure of argumentation are analyzed in works [4,5], which propose approaches for formalizing a logical-semantic model of text aimed at evaluating the soundness of scientific claims. The application of these ideas in domain-specific computational modeling is also supported in [6], where the use of tropes and semantic spaces is suggested for refining concepts within models.

Technical aspects of automatically deriving ontologies from texts are addressed in [7], which describes the capabilities and limitations of ontology learning methods from text, especially relevant in the case of scientific publications.

A separate group of studies focuses on the application of large language models (LLMs), particularly GPT, for generating reviews and evaluating scientific articles. In [8], the use of GPT for preliminary reviewing is proposed, though without clear logical-structural control over the text. A similar approach is discussed in [9], which demonstrates that LLMs can produce stylistically convincing, yet often superficial reviews.

The evaluation of factual consistency of claims in scientific publications is implemented in [10], where it is verified whether a specific claim is supported by data. However, this model does not assess the structure of scientific discourse as a coherent argumentative entity. An extended study of inconsistency risks in language models such as GPT-3 was conducted in [11], which highlights potential errors in generating complex logical dependencies. In [12], the possibility of using language models to reduce information asymmetry is discussed, yet these approaches do not operate with a structured representation of knowledge and do not include logical-discursive validation of argumentation at the level of a scientific publication, as proposed in our system.

The use of automatic indices and services for analyzing scientific output is also well known, such as Microsoft Academic Service [13], which constructs scientific knowledge graphs but does not perform logical validation of argumentative coherence.

Institutional issues of scientific publication quality are reflected in the lists of discontinued Scopus journals [14], where the reasons for discontinuation are often associated with the absence of proper peer review or with the formal character of publications. An additional ethical guideline is provided by the COPE recommendations [15], which emphasize the necessity of logical consistency and structural completeness in the creation and reviewing of scientific texts.

In this context, our study continues the line of research combining ontologies and language models, but for the first time implements them as a formalized logic-ontological structure applied to the full text corpus for the purpose of detecting structural deficiencies. This makes it possible not only to assess factual accuracy or style, but also to formally verify the presence and coherence of elements of scientific argumentation.

3 MATERIALS AND METHODS

In this study, the object of analysis is a scientific article composed of textual fragments corresponding to the main logical-discursive components: hypothesis, method, results, conclusions, and so forth. The expected outcome is a structured representation of the article in the form of a knowledge graph, which includes formalized nodes and links between them according to the constructed ontology, as well as an analytical assessment of the integrity, completeness, and logical coherence of the scientific argumentation, expressed through a set of quantitative metrics.

The methodology is based on the principle of multi-level system integration of three interrelated components: a formal ontology, logical-discursive verification, and a large language model (LLM). The formal ontology describes the semantic structure of the scientific text in the form of a set of terms and admissible relations among them. The verification logic defines a set of rules that must be fulfilled for the publication to be considered structurally complete. The LLM provides semantic alignment between text fragments and ontological types, as well as models the logical dependencies among them. In this way, the language model functions as an execution mechanism that performs structural analysis and logical validation in accordance with the formal problem specification.

The overall workflow of the study is implemented as a sequence of interconnected stages. At the first stage, the publication text undergoes preprocessing: segmentation into logical fragments (introduction, methods, results, etc.) is performed. At the second stage, each fragment is assigned an appropriate ontological type using the language model and predefined instructions. Then, an ontological knowledge graph is constructed, in which nodes correspond to fragments and directed edges represent logical relations (e.g., “method validates hypothesis”, “conclusion is based on results”). At the next stage, the graph is checked against a set of logical-discursive rules implemented as Boolean predicates. After this, the system computes an aggregated score of logical and structural coherence. The final stage is the generation of a textual review

DOI 10.15588/1607-3274-2025-4-7

based on the identified nodes, assessed relations, and detected violations, integrated into a template of an automated report.

This approach ensures control over the structure of the study, reproducibility of the reviewing process, and the potential scalability of the system to other domains of knowledge or types of texts.

Ontology, in the context of this study, is defined as a formal, explicit, and machine-processable specification of a set of concepts and the relationships among them, which describes the structure of knowledge within a specific subject domain. Unlike a taxonomy or glossary, an ontology defines not only a hierarchy of concepts, but also the types of connections, logical dependencies, and constraints that operate between conceptual units.

The goal of constructing an ontology of a scientific publication is to formalize the logical-discursive structure of the text, that is, to identify the key semantic nodes that are necessarily present in a scientific argument and to fix the normative relationships among them. The ontology is implemented as a set of node types (terms) and admissible relations between them, forming a directed semantic graph. A typical set of terms includes such ontological nodes as Hypothesis, Goal, Method, Object, Result, Conclusion, Novelty, Practical significance, and so forth. In the future, individual nodes may be refined with additional attributes (e.g., structure, formal features of formulation, stylistic characteristics, etc.).

Admissible relationships between types of ontological nodes are defined by a set $\Re$, which is explicitly specified as a list of tuples of the form $(t_i, t_j, type)$. Each element $r \in \Re$ describes a permitted normative relation between a node of type t_i and a node of type t_j, with the type of connection indicated. For example, the tuple (“Method”, “Hypothesis”, “validates”) is interpreted as “the method validates the hypothesis”, while (“Conclusion”, “Result”, “justified_by”) means “the conclusion is justified by the results.”

The types of relations are formalized as predicates such as `validates`, `supports`, `derived_from`, `motivates`, `contradicts`, and others. Each of these predicates takes two arguments: the first is the source node, and the second is the target node. For example, the predicate `validates(Method,Hypothesis)` indicates that the method validates the hypothesis, while `supports(Result, Hypothesis)` means that the result supports the hypothesis.

The set $\Re$ is part of the input ontological model provided to the agent along with the set of predicates Φ used for logical-discursive validation. This set is editable and can be adapted to the requirements of a particular journal or research paradigm. It defines the expected structure of scientific argumentation, according to which knowledge graphs are constructed and semantic evaluation of texts is performed.

The developed ontology is not universal – it is tailored to the format of academic articles, the IMRaD structure, and the recommendations of COPE. During the construction of the ontology, elements from foundational knowledge ontologies (such as BFO and UFO) were employed alongside specialized categories proposed within our model. A distinctive feature of the developed ontology is the inclusion of additional concepts that are typically not represented in the formal structure of scientific articles but are important for peer review. These include features such as scientific relevance, alignment with contemporary scientific discourse, linguistic neutrality, stylistic correctness, and the representativeness of the reference base.

Each ontology node is associated with a fragment of the article text. To achieve this, a labeling function $L: V \to T$ is introduced, which maps a node v_i from V (extracted fragments) to a type from the set T. All text fragments that are not assigned to a label are considered out-of-scope or noise. The construction of the ontological graph $G = (V, E, L)$ is performed as a result of text interpretation by the language model, which conducts semantic classification of fragments to define the labeling function L and establishes admissible links E according to the specified rules $\Re$.

The model allows for incompleteness: not all nodes may be detected; however, the presence of a certain minimal core is a prerequisite for the formal eligibility of a text for review. Thus, ontology serves as a structural scaffold that defines the expected architecture of scientific discourse and provides a basis for logical-structural validation.

Classical knowledge ontologies such as BFO and UFO are primarily oriented toward modeling universal types of entities and their properties in the real or conceptual world. Their advantages lie in a high level of abstraction, ontological rigor, and stability. However, when modeling the logical-discursive structure of a scientific text, this level of abstraction is insufficient to capture the dynamics of argumentation, semantic dependencies between text fragments, and elements of publication quality assessment.

Within the scope of our study, several essential extensions to the base ontological schema were introduced. First, dedicated nodes were added for concepts that are not structurally mandatory in a paper but are critically important for peer review: Scientific relevance, Practical significance, Embeddedness in contemporary scientific discourse, Novelty, Completeness of justification, and Interpretation of results. These nodes enable the system to track not only the presence of standard elements (hypothesis, method, result) but also the degree of their substantiation and interconnectedness.

Second, our focus shifts from the ontological hierarchy of types to the dynamics of relationships between them. Instead of classical typing with subclasses and object properties, we use a model of logical predicates over node pairs that capture discursive logic: for example, `supports(H, R)`, `derived_from(M, H)`, `justified_by(C, R)`. In this way, the ontology em-

DOI 10.15588/1607-3274-2025-4-7

phasizes functional relations that reflect the logic of scientific argument construction.

Third, meta-concepts have been introduced, allowing not only structural elements but also their quality to be described. For instance, a node of type *Hypothesis* may have attributes such as explicit, testable, narrow vs broad. This makes it possible to combine structural analysis with content-based criteria evaluation.

These extensions exemplify the adaptation of foundational ontologies to the applied task of automated analysis of scientific texts for the purpose of assessing their logical-discursive completeness. The proposed modifications do not contradict the foundations of BFO or UFO but rather function as applied overlays tailored to the needs of modern scholarly communication and automated reviewing.

Unlike the general ontology of a scientific publication, which captures the structure of scientific discourse as such, the reviewing ontology formalizes the requirements for its individual components and defines criteria for logical validation of their presence, interconnection, and semantic completeness. It is used not to describe the article, but to evaluate it in accordance with expert standards of academic quality.

At the first level, such ontology contains a specification of requirements for each key node of a scientific publication. For the Hypothesis, the requirements include clarity of formulation, testability, connection to a problem situation, and justification of relevance. The Method must be related to the hypothesis, described in sufficient detail to enable reproducibility, and include a rationale for its selection. The Results must be obtained through the application of the method and presented in a form suitable determines whether the conclusions are connected to the results through an admissible relation of type `justified_by`.

In the implemented system, the set of predicates $\Phi = \{\phi_1, \ldots, \phi_k\}$ is defined as an explicit collection of Boolean conditions formulated based on the ontological model (node types and admissible links) and the structure of the review template (logical- for analysis. The Conclusions should follow from the results, generalize them, and not merely restate the hypothesis verbatim.

In addition to these core nodes, the system separately checks for Novelty, Practical significance (applicability), Justification (sufficient argumentation), and Scientific relevance (connection to the current state of the problem in the global context).

All these requirements are formalized as a set of Boolean predicates:

$$\Phi = \{\phi_1, \phi_2, \ldots, \phi_k\},\ \phi_i : G \to \{0,1\},$$

where each ϕ_{i} is a logical-discursive rule that verifies whether a specific condition is met within the ontological graph GGG. For example, the rule $\phi_{\text{hypothesis\_present}}$ checks for the presence of at least one node labeled with the type Hypothesis; the rule $\phi_{\text{method\_mathces\_hypothesis}}$ verifies whether a logical connection of type `validates` exists between the method and the hypothesis; and the rule $\phi_{\text{result\_supports\_conclusion}}$ discursive requirements for the scientific text). Each rule ϕ_{i} checks the fulfillment of a specific condition in the graph $G = (V, E, L)$ – for example, the presence of a node of type Hypothesis, or a link of type `validates` between the method and the hypothesis. At the same time, the system supports the possibility of dynamically extending the set of predicates, considering the editorial policy of a journal, domain-specific features, or individual review settings. This enables flexible adaptation of the logical-discursive validation to the context of the submitted publication, while preserving the transparency and formalization of the process.

A key stage is internal self-verification: prior to generating the review, the agent explicitly computes the value of each rule $\phi_i(G)$ and creates a logical mask of violations. For each violated rule $\phi_i(G) = 0$, an explanation is generated, which is subsequently integrated into the corresponding section of the review. In this way, a link is established between structural validation and natural language justification, enhancing the interpretability of the system.

A violation of any rule from the set Φ is interpreted as a specific anomaly in the structure of the scientific text, indicating logical incompleteness, lack of argumentative coherence, or a missing key component of the publication. The totality of such violations forms the set:

$$\Delta = \{\phi_i \in \Phi \mid \phi_i(G) = 0\},$$

which represents a list of specific logical-discursive deficiencies. This set serves as the basis for constructing an interpretable review or a formalized analytical report.

The overall measure of a manuscript's compliance with the requirements of the reviewing ontology is defined as the proportion of fulfilled rules:

$$\text{Consistency}(G) = \frac{|\Phi| - |\Delta|}{|\Phi|},$$

where $|\Phi|$ is the total number of rules, and $|\Delta|$ is the number of violations. This metric is used as a numerical indicator of structural completeness and logical consistency.

Thus, the agent not only generates the text of the review but also performs a formalized logic-ontological analysis based on a normative model of scientific discourse. The proposed approach allows for the validation of both basic structural requirements (such as the presence of key elements) and more complex logical dependencies between them, for instance, the alignment of the method with the hypothesis, the justification of conclusions by results, or the substitution of scientific novelty with practical applicability.

The list of predicates is provided in the appendix as a table, representing the core of the set Φ – that is, the pri-

DOI 10.15588/1607-3274-2025-4-7

mary set of Boolean conditions used to assess the structural completeness and logical consistency of a scientific text. This core is passed to the agent as part of the input ontological structure for logic-discursive analysis. If necessary, the list of predicates can be extended or adapted according to the specific subject domain or the editorial policy of a particular scientific journal. Based on the results of predicate evaluation, the system generates a structured report that reflects the degree to which the publication satisfies the expected criteria of argumentativeness, completeness, and coherence of scientific reasoning.

The formalization of a scientific article's text is carried out through the construction of a semantic knowledge graph that models the logical-discursive structure of the publication. The source text S is treated as an ordered set of fragments s_i, which may contain hypotheses, methods, results, conclusions, and other components of scientific argumentation. These fragments are mapped to nodes of the graph $G=(V,E,L)$, which serves as the primary object of analysis. The graph $G=(V,E,L)$ is automatically constructed by the agent based on the article text and the ontological model, which includes the set of terms S and the set of admissible links $\Re$. The graph's nodes are created via semantic classification of text fragments, while the edges are formed by checking for the presence of discursive dependencies according to the ontology.

In the implemented system, the resulting graph can be exported in a standard format (e.g., .json or .dot) or visualized. This enables transparent inspection of the structure G, in particular – the presence of all key nodes, the types of relations, and the possibility of evaluating the logic-discursive predicates. The graph G, serves as the foundation both for structural analysis and for the generation of the review.

The set $V=\{v_1,v_2,\ldots,v_n\}$ is the set of nodes in the graph, where each node v_i corresponds to a specific text fragment s_i. Each node is assigned an ontological type using a labeling function:

$$L:V\to T,$$

where T is the set of ontological terms (such as Hypothesis, Method, Result, Conclusion, Novelty, Significance, etc.). This labeling is performed automatically based on text analysis using a GPT-based language model, which generates the labeling function LLL through semantic alignment of fragments with the definitions of ontological nodes.

The set $E\subseteq V\times V$ defines the logical connections between nodes. Each edge $(v_i,v_j)\in E$ represents a relation of discursive dependency, for example, "the method validates the hypothesis" or "the conclusion is based on the result". The type of relation is specified by a separate function that assigns predicate-level semantics to the link, according to the reviewing ontology. Thus, the graph GGG is not merely a textual schema but a structured representation of the argumentative logic of the publication.

The formalized graph enables verification of the presence of all required structural components and the logical dependencies among them. For this purpose, the previously defined numerical model of structural consistency – Consistency Score – is used.

Thanks to the construction of the knowledge graph G, it becomes possible not only to detect formal deficiencies in the text but also to perform automated generation of a review based on the structure of the graph, its compliance with the expected logic, and the violations identified. This approach allows for a shift from surface-level analysis of style and content to formalized semantic verification of the scientific text.

The construction of the knowledge graph G plays a central role in the architecture of the proposed system: it provides a formal transition from the linguistic representation of the text to a structured ontological model suitable for logical-discursive analysis. The graph structure makes it possible not only to record the presence or absence of key elements but also to verify their logical consistency. This, in turn, serves as the basis for computing the Consistency metric and generating an explainable review. Thus, formalization in the form of a graph transforms the evaluation of a scientific text from a heuristic procedure into a formalized and replicable operation.

The constructed graph $G=(V,E,L)$, in this context, is not merely a knowledge graph in the general sense but a specialized ontological graph, since all its elements – both nodes and edges – are formed in accordance with a predefined ontology. Specifically, the node types are restricted to a set of terms T (e.g., Hypothesis, Method, Result), and the links are limited to a set of admissible relations $\Re$. This representation not only preserves the semantic structure of the text but also enables validation of its compliance with expected scientific standards using the set of predicates Φ.

To validate the developed logic-ontological model for analyzing scientific publications, a specialized test corpus was constructed, including texts with deliberately introduced logical defects. This allows for both quantitative and qualitative evaluation of the system on texts for which the violated structural or logical elements are known in advance and thus provides a basis for assessing the model's ability to detect them.

The first part of the corpus consists of artificially generated texts in which one or more structural components were deliberately omitted or altered. Specifically, texts were created without a hypothesis, without a method, with results that do not support the hypothesis, with conclusions that do not generalize the results, with confusion between practical significance and scientific novelty, and with argumentation limited to local context without properly addressing the broader scientific discourse. These texts maintain syntactic correctness, an academic-like style, and a length typical of short articles. They allow precise evaluation of whether the system can detect the

DOI 10.15588/1607-3274-2025-4-7

absence of key nodes, logical inconsistencies, or argumentation defects without confusing them with stylistic variation.

The second part of the corpus includes real-world publications from journals that were excluded from reputable international databases due to systematic violations of academic integrity. These publications are treated as a heuristic negative class – a source of texts that may appear formally scientific but typically contain violations of logical-discursive structure, lack of justification, missing hypotheses, misuse of methods, or overgeneralized conclusions.

The initial documents were obtained in PDF and DOCX formats, then converted to plain text to ensure uniformity of the input data. The texts were further preprocessed to standardize formatting, remove unnecessary breaks, and conform to a unified structure. For each document, a manual or semi-automated annotation was created in JSON format, which includes information about the expected structural nodes, their presence or absence, and known logical-discursive defects. These annotations serve as a reference for computing error detection accuracy and validating the model.

Thus, the constructed corpus enables comprehensive testing of the proposed approach using both real and synthetic examples with varying levels of structural and argumentative quality. It also ensures reproducibility of the experiment and independent verification of the model's effectiveness on external material.

To evaluate the effectiveness of the proposed ontology-driven approach to scientific publication analysis, a set of quantitative metrics was used to assess how accurately the system identifies structural nodes, logical dependencies between them, and generates adequate reviews. These metrics incorporate both classical information retrieval indicators and specialized metrics for ontological analysis.

Several key metrics are used to assess the quality of the system's performance. In particular, precision is defined as the proportion of correctly identified defects among all defects flagged by the system. It is calculated using the formula:

$$\text{Presision} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FP}|},$$

where $|\text{TP}|$ is the number of true positives (correctly detected violations), and $|\text{FP}|$ is the number of false positives.

Recall characterizes the system's ability to detect all existing defects and is defined as the proportion of detected defects among all those that are actually present:

$$\text{Recall} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FN}|},$$

where $|\text{FN}|$ is the number of defects that the system failed to detect.

Another important metric, in addition to the Consistency Score, is the Onto-score, which reflects the system's ability to correctly identify key ontological nodes (such as Hypothesis, Method, Result, Conclusion, etc.). This metric is computed as the ratio of correctly identified nodes to the total number of expected nodes:

$$\text{OntoScore} = \frac{|V_{\text{correct}}|}{|V_{\text{expected}}|}.$$

For each test article, the automatically constructed ontological graph is compared with a reference structural model that defines the expected presence and types of nodes according to the ontology. Compliance is assessed by comparing node types (for the Onto-score) and by evaluating the fulfillment or violation of logical rules (for the Consistency metric).

The results of the experiments are recorded in tabular format, where each row corresponds to an article and includes: its identifier, the number of detected nodes, the number of violated rules, and the values of the Precision, Recall, Consistency, and Onto-score metrics. For visualization, bar charts are used (to compare Precision and Recall across articles), heatmaps are used to show the frequency of detection of specific defect types, and pie charts provide an overview of the review results.

Thus, the result verification process is quantitative in nature, ensures reproducibility, and makes it possible to clearly identify the strengths and weaknesses of the ontology-driven reviewing system.

4 EXPERIMENTS

The objective of the experiment was to test the hypothesis that the implemented system is capable of automatically detecting structural and logical-discursive defects in scientific articles based on an ontological knowledge graph, without human intervention.

For each article, the following sequential steps were performed:

(1) loading and cleaning the text;

(2) segmenting the text into logical fragments;

(3) semantically labeling each fragment with an ontological node type using the publication ontology via the GPT agent;

(4) constructing the graph $G = (V, E, L)$;

(5) verifying compliance with the set of predicates Φ;

(6) calculating the metrics: Precision, Recall, Consistency, and Onto-score;

(7) automatically generating a draft review.

The processing steps were implemented as sequential functions of the GPT agent, aligned with the logic of distinct modules: semantic labeling, graph construction, rule validation, and review generation.

In total, 120 texts were analyzed, divided into three subsets of 40 texts each:

A – artificially generated texts with controlled defects (removed nodes, broken links);

DOI 10.15588/1607-3274-2025-4-7

B – publications from journals excluded from scientometric databases (formally structured but often lacking novelty or argumentative rigor);

C – benchmark articles selected from open-access publications in Q1-category journals indexed in the Scopus scientometric database.

All publications were processed using a single ontological JSON template, in which the set of node types, admissible relations, and validation rules were manually defined. This ensured a unified analysis structure and consistent comparison of results.

The experiments were conducted in the Jupyter Notebook environment (Python 3.10). For graph construction, the networkx library was used; for table processing, pandas; for metrics calculation, scikit-learn; and for visualization, matplotlib and seaborn. The GPT agent was implemented on the ChatGPT GPTs platform (without using an external API), utilizing the ontological structures and review template as initial context.

Validation of the graph's compliance with the reviewing ontology requirements was performed through the logic-ontological core – a module that implements a set of Boolean predicates $\phi_i : G \to \{0,1\}$ to evaluate the satisfaction of each structural condition.

For each article, a table was generated showing the number of nodes, fulfilled predicates, and computed metrics. The following visualizations were produced:

(1) bar charts for Precision and Recall across all classes;

(2) boxplots to compare metrics between classes;

(3) heatmaps of rule fulfillment across all documents;

(4) sample ontological graphs for individual articles.

This enabled assessment of the system's quality, robustness, and sensitivity to different types of defects.

5 RESULTS

To conduct the experiments, a corpus was constructed consisting of three subsets of scientific publications, differing in structural completeness and the expected quality of logical-discursive content. General statistics are presented in Table 1. Each subset contains the same number of documents (40), ensuring symmetry for comparative analysis. The average number of extracted fragments (nodes) in documents from subset C was 7.6, corresponding to the typical structure of a full-length scientific article. In subset B, this value was lower – 5.1, and in subset A – 4.2, indicating the presence of structural omissions.

For each subset, the detection frequency of four key ontological nodes was evaluated: Hypothesis, Method, Result, and Conclusion. As shown in Figure 1, in the benchmark articles (subset C), these nodes were identified with high consistency (in 95–100% of cases). In subset B, a significant reduction was observed in the presence of Conclusions (down to 60%) and Methods (down to 55%), whereas in subset A, where the defects were deliberately introduced, the frequency of specific node types could be below 40%

The obtained data reveal significant differences in structural completeness among the article classes, providing a foundation for further analysis of the effectiveness of logic-ontological verification and the calculation of quality metrics.

One of the key stages of the system's operation is the automatic semantic labeling of text fragments using a large language model in accordance with the ontological schema. To evaluate the quality of this labeling, precision and recall were calculated for each of the main node types: Hypothesis, Method, Result, and Conclusion.

Table 1 – Distribution of Documents Across Corpus Subsets

Parameter	A (Synthetic)	B (Discontinued Jounals)	C (Benchmark)
Number of documents	40	40	40
Average number of nodes	4.2	5.1	7.6
Standard deviation	1.1	1.4	0.9

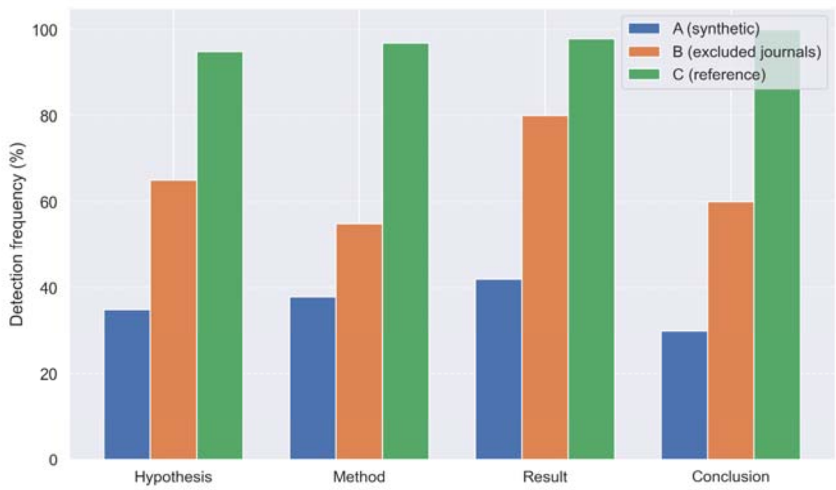

Figure 1 – Frequency of Detection of Core Ontological Nodes Across Corpus Subsets

DOI 10.15588/1607-3274-2025-4-7

In the benchmark article class (C), the labeling precision exceeds 0.90 for all four node types, indicating the model's ability to correctly identify the discursive structure of full-format scientific texts. In contrast, subset B shows a decrease in both precision and recall, especially for Result (Precision = 0.74, Recall = 0.68) and Conclusion (Precision = 0.70, Recall = 0.64). The lowest scores were observed in the artificially deformed texts (subset A), where, for example, conclusion elements were deliberately omitted or replaced with general statements. Here, the precision for Conclusion was only 0.48, and recall – 0.37.

The labeling quality evaluation is presented in Table 2. The values were computed based on comparison with the reference annotation, which included manual labeling of all expected nodes.

Table 2 – Precision and Recall Metrics for Ontological Categories (Averaged Across Subsets)

Node	A		B		C	
	Precision	Recall	Precision	Recall	Precision	Recall
Hypothesis	0.61	0.58	0.78	0.75	0.93	0.91
Method	0.67	0.63	0.80	0.78	0.94	0.92
Result	0.54	0.49	0.74	0.68	0.91	0.89
Conclusion	0.48	0.37	0.70	0.64	0.92	0.90

Examples of constructed knowledge graphs based on semantic labeling are presented in Fig. 2. The left part illustrates the structure of a benchmark article, in which all key ontological nodes are present: Goal, Hypothesis, Method, Result, Conclusion, and the relations between them are clearly defined. The right part shows the structure of an artificially generated publication in which the Result node was deliberately omitted. As a result, the graph appears fragmented and lacks a logical transition to the conclusion, which was accurately detected by the system.

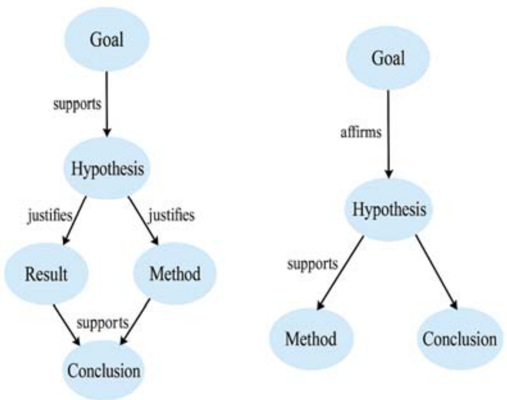

Figure 2 – Knowledge Graphs for a Benchmark Article (left) and a Synthetic Article (right)

These results indicate the model's sufficient capability to automatically recognize the key structural elements of a scientific article. At the same time, the analysis shows that the labeling process is sensitive to the quality of the input text and can detect deviations even when the discursive structure is weakly expressed.

The next stage of evaluating the effectiveness of the ontology-driven model was to determine the degree of compliance of the knowledge graph with the discursive rules defined by the reviewing ontology. To this end, each constructed graph $G = (V, E, L)$ was checked against a set of rules $\Phi = \{\phi_1, \phi_2, \ldots, \phi_k\}$, which represent the expected relations between key nodes.

The results revealed a significant difference across the three subsets. In class C (benchmark articles), the average proportion of satisfied rules was 92%, in class B (problematic real-world articles) 71%, and in class A (artificially defective) – only 55%. Thus, the Consistency index reflects the degree of alignment between the internal logic of an article and the formal standards of scientific exposition.

The most frequently violated rules were:

– Method validates Hypothesis – absence of or weak linkage between the method and the hypothesis;

– Conclusion justified by Result – formal conclusions that are not based on the results;

– Hypothesis motivated by Literature Review – hypotheses not substantiated in the introduction or literature review.

Fig. 3 presents a heatmap showing the frequency of rule satisfaction across the different subsets. Each row corresponds to a specific discursive rule, and each column – to a class of articles. The visualization clearly shows consistent rule fulfillment in the control group, whereas synthetic defects systematically disrupt basic relations. Class B exhibits a more heterogeneous pattern: some articles are structurally complete, while others demonstrate significant logical deficiencies.

The obtained results demonstrate the sensitivity of the ontological model to implicit violations in the logic of scientific exposition. Even when the required fragments are formally present, the system is able to detect broken or missing links between them – an essential capability for preliminary review and recommendations for revision.

To quantitatively assess structural completeness and compliance with the logical-discursive structure, two integral metrics were used: Consistency and Onto-score. The first reflects the proportion of satisfied discursive rules from the set Φ within the constructed ontological graph, while the second indicates the proportion of correctly identified structural nodes relative to the expected reference set. In subset C (benchmark articles), the average Consistency score was 0.92 and the Onto-score was 0.94, indicating a stable structure and full presence of key discursive components. In subset B, both metrics showed

DOI 10.15588/1607-3274-2025-4-7

a noticeable decline: 0.71 for Consistency and 0.76 for Onto-score. The lowest values were recorded in subset A, where due to controlled defects the metrics did not exceed 0.56 for Consistency and 0.61 for Onto-score.

Fig. 4 presents a comparison of the distributions of Consistency and Onto-score metrics across the three classes using boxplot diagrams. It can be observed that benchmark articles exhibit not only high average values but also low dispersion, while subset B demonstrates a widespread, indicating varying degrees of structural completeness even within the same class.

In addition, Fig. 5 presents histograms of the Consistency score distribution for all documents in the corpus. This type of visualization makes it possible to assess not only the average but also the distribution density of the metric values within each class. In subset A, most values are concentrated in the interval 0.4-0.6, whereas in class C, the majority of articles exhibit Consistency scores above 0.9.

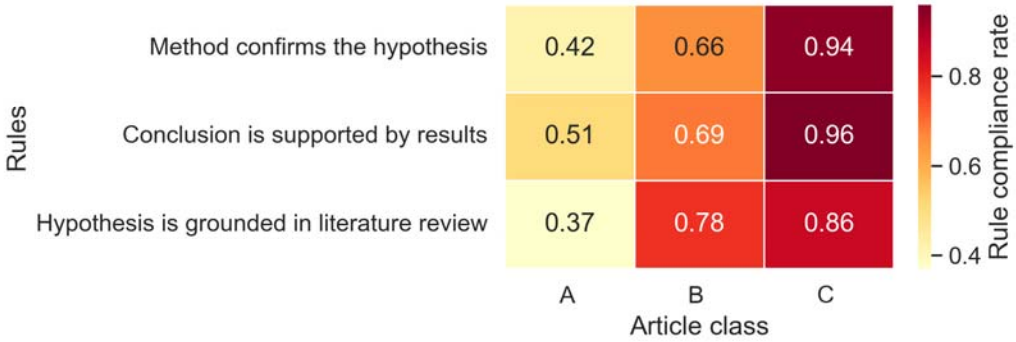

Figure 3 – Heatmap of Logic-Discursive Rule Satisfaction Across the Three Corpus Subsets

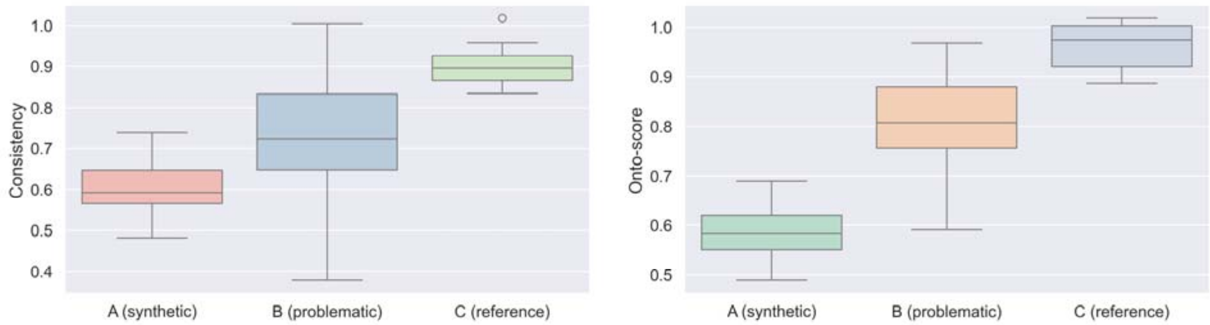

Figure 4 – Boxplots for the Consistency and Onto-score metrics across three subsets

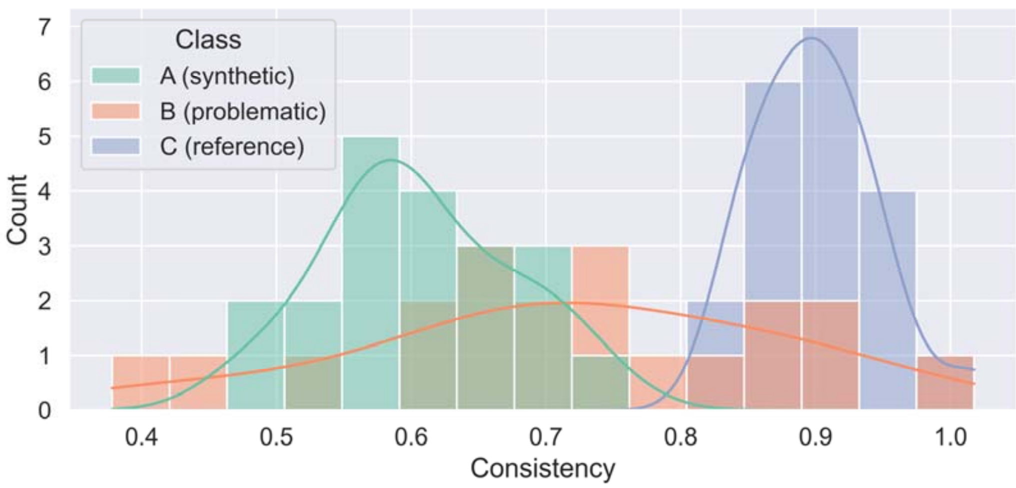

Figure 5 – Histogram of Consistency Score Distribution for All Corpus Articles

DOI 10.15588/1607-3274-2025-4-7

The obtained results confirm that both metrics are informative criteria for the preliminary automated assessment of the structural quality of scientific publications. They effectively distinguish between text classes and can serve as quality indicators even in the absence of expert input.

To evaluate the system's ability to accurately detect structural defects in scientific publications, classical information retrieval metrics – precision and recall – were employed. Precision indicates the proportion of truly detected violations among all automatically flagged ones, while recall characterizes the system's ability to identify all defects actually present in the text according to the reference annotation.

In subset A, where the defects were deliberately embedded in the text, the system achieved a precision of 0.82 and recall of 0.79, demonstrating its ability to detect even subtly masked structural issues. In subset B, which contains real articles with potentially poor structure, the values were lower – 0.74 for precision and 0.66 for recall. In the control class C, the system detected only minor deviations – mostly false positives – leading to a reduction in precision to 0.71, while recall remained high at 0.93.

Fig. 6 presents a bar chart comparing the average values of Precision and Recall across each subset. As shown, subsets A and B exhibit similar trends, although the defects in subset A are detected more consistently, due to their formalized nature. In subset C, Recall dominates, reflecting a scenario in which the system tends to "overreact," responding even to stylistic deviations. Edge cases were analyzed separately.

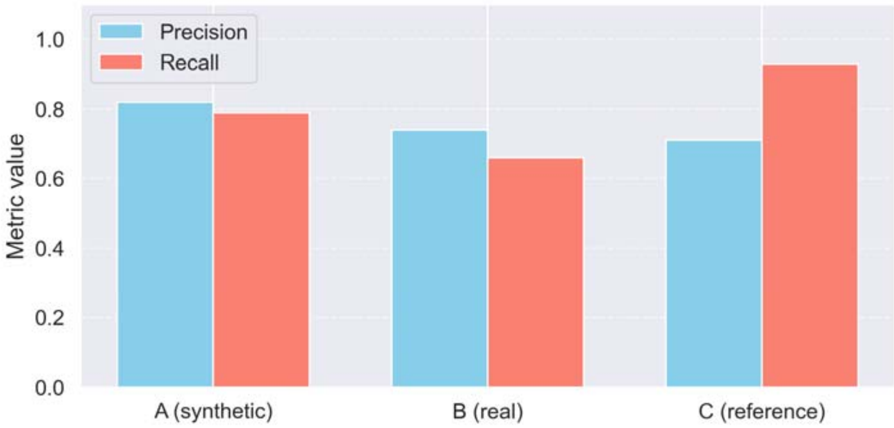

Figure 6 – Comparison of Precision and Recall Across Subsets

The highest precision score 1.0 was recorded for a synthetic article from subset A, in which three critically important nodes (Method, Result, Conclusion) were simultaneously removed. This allowed the system to clearly detect the defects without generating false positives. Conversely, the lowest precision score 0.54 was observed in a real article from subset B, where the system mistakenly interpreted general phrases as structural nodes, resulting in a large number of false positives.

Thus, the use of Precision and Recall metrics not only allows for the evaluation of overall defect detection performance but also reveals specific scenarios in which the model is either too conservative or overly "sensitive" to semantic variation.

To gain deeper insight into the typical types of defects detected by the system, a pie chart was constructed (Fig. 7) showing the distribution of missing ontological nodes in subset A – that is, in the artificially deformed texts. The chart complements the Recall metric: each segment represents the proportion of missing nodes for a given ontological type.

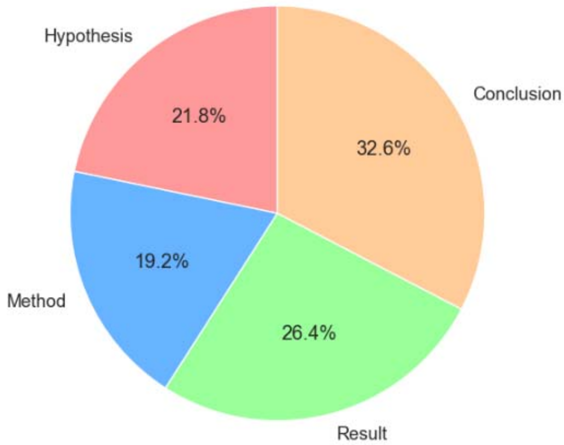

Figure 7 – Distribution of Missing Ontological Nodes in Subset A (Based on Recall Metric)

As shown in the diagram, the largest proportion of missing nodes corresponds to the Conclusion type (63%), indicating the difficulty in identifying logical closures in deformed or overly generalized statements. This is followed by Result (51%) and Hypothesis (42%), which points to challenges in detecting research justification and outcome representation. The fewest losses were recorded

DOI 10.15588/1607-3274-2025-4-7

for the Method type (37%), which aligns with the fact that methodological descriptions often have a distinct stylistic structure.

This visualization allows not only for the quantitative assessment of Recall but also for the localization of the most problematic areas in article structures, which is important for further improvement of the recognition model.

To qualitatively illustrate the system's operation, typical examples of knowledge graphs are provided, which in this context are constructed as ontological graphs. Each such graph $G = (V, E, L)$ is built based on the article text and a predefined ontology. It reflects the logical-discursive structure of the article: the nodes V correspond to semantic fragments, the labeling function L assigns them ontological categories, and the edge set E forms the normative logic of argumentation.

This subsection presents one graph for each of the corpus subsets (A, B, C), enabling visual comparison of typical deviations – from structural violations to logical incompleteness and full compliance.

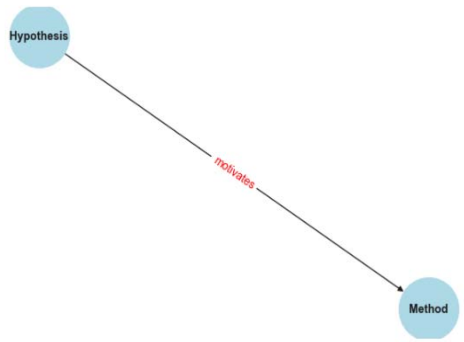

Figure 8 – Ontological Graph for an Article with Controlled Structural Defects (Subset A)

Fig. 8 shows an example graph for an article from subset A, in which the Result and Conclusion nodes were deliberately omitted. As a result, the Hypothesis and Method remained without logical closure: relations of type `derived_from` and `justified_by` are missing. The graph appears fragmented, with a broken argumentative chain, which was correctly detected during logic-discursive validation.

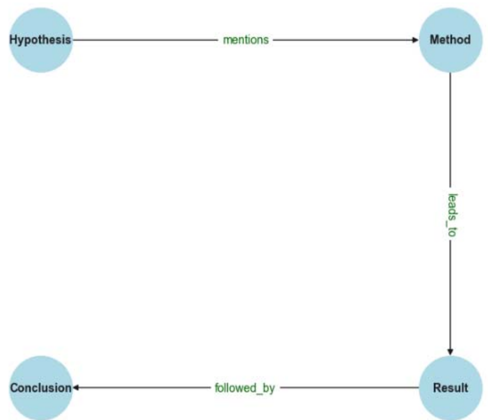

Figure 9 – Graph for a Problematic Article with Violations in Logical Structure (Subset B)

Fig. 9 illustrates an example graph for a publication from subset B. Unlike the previous case, all four key nodes are present here; however, their logical coherence is violated. In particular, the method is not specific to the hypothesis, and the results are not connected to the conclusions. As a result, the graph is formally complete but logically flawed: the `validates` and `justified_by` edges are either missing or lack meaningful justification in the text.

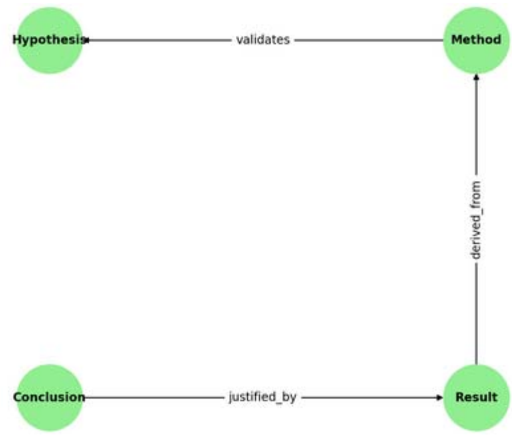

Figure 10 – Fully Coherent Ontological Graph for a High-Quality Article (Subset C)

Fig. 10 presents a sample graph for a benchmark publication from subset C. In this case, all major nodes and the logical links between them are present: Method validates Hypothesis, Results are derived from the method, and Conclusions provide a justified summary of the results. This structure is logically complete and corresponds to the expected model of scientific exposition.

Although the graphs for subsets B and C may appear visually identical – with the same nodes (Hypothesis, Method, Result, Conclusion), the same number of edges, and identical geometric layout – their fundamental difference lies in the semantics of the edges. In graph C, which corresponds to a benchmark publication, all connections are normative: `Method` validates `Hypothesis`, `Result` is derived from `Result`, and `Conclusion` is logically justified by `Result`. In contrast, in graph B, these same connections are replaced by stylistically neutral or semantically weak relations (`mentions`, `leads_to`, `followed_by`) which deprive the graph of logical force and render it formally complete but conceptually flawed.

This difference in the types of relations allows the system to detect violations of normative structure, even when the external form of the graph remains unchanged.

The comparison of graphs across subsets makes it possible to visually identify key differences: from missing nodes (subset A), through inconsistent relations (subset B), to logical completeness (subset C). The graph-based representation not only formalizes the scientific text but also serves as a powerful tool for the automated diagnosis of its structural quality.

DOI 10.15588/1607-3274-2025-4-7

6 DISCUSSION

The results obtained confirm that the use of an ontological model combined with logic-discursive validation enables effective detection of structural and semantic defects in scientific articles prior to the peer review stage. High Consistency and Onto-score values in the control group demonstrate that the system correctly identifies complete scientific argument structures, and in cases with intentionally disrupted or implicitly flawed logic, it is capable of detecting missing or inconsistent key components. This highlights not only the technical feasibility of the proposed approach but also its conceptual productivity.

Comparison with related research, particularly with approaches that rely on rule-based or formally structural models without deep semantic interpretation, indicates the advantage of a hybrid methodology that combines ontological reasoning, logical verification, and the generative capabilities of LLMs. Unlike traditional systems based on template-driven segment detection, our model operates at the level of semantic relations and context-sensitive logic of exposition. On the other hand, the presence of false positives in articles from the benchmark subset reveals the sensitivity limits of the model, consistent with findings that emphasize the constraints of LLMs when dealing with highly variable textual forms.

The similarity of results in subsets A and B, despite the different nature of the defects (artificially introduced vs. naturally occurring), indicates that the model is equally capable of responding to both explicit and latent deviations from normative structure. At the same time, in subset B, the system encounters more complex error configurations, explaining the wider spread of metrics and the higher proportion of logically ambiguous cases.

Among the limitations, it is worth noting the model's sensitivity to stylistic variations and non-formalized content presentation. In cases where the author deliberately avoids strict academic structure (e.g., does not explicitly mark the hypothesis or merges conclusions with discussion), the system may lower the evaluation score despite the presence of valid scientific content. This highlights the need for further extension of the ontology to encompass discursive strategies.

The practical applications of the proposed approach include automated support for initial editorial screening, review systems with low expert capacity, and educational platforms for teaching scientific writing. Integration of the model into the reviewer interface enables quick identification of weak points, structured recommendations, and enhanced transparency of the evaluation process.

Promising directions for future research include improving the logical core of the system using formal logics (e.g., description logics), leveraging vector-based ontologies for assessing semantic proximity of nodes, and testing the system on materials from diverse scientific domains. A separate objective is the creation of an open-access corpus of articles with logic-ontological annotation, which would enable standardization and advancement of metareviewing methodologies.

CONCLUSIONS

In this work, the scientific task of formalized automated assessment of the structural quality of scientific publications was addressed by developing and applying an ontology-driven model of discursive analysis. The proposed approach integrates the ontological formalization of a scientific publication, logic-discursive consistency verification, and semantic labeling using a large language model.

The scientific novelty of the research lies in the following. For the first time, an integration of the ontology of a scientific publication and the ontology of peer review is proposed to construct a formalized knowledge graph $G=(V,E,L)$ that captures both the structure and logic of an article. The logic-discursive verification approach has been further developed by defining a set of rules Φ that formalize the requirements of scientific exposition and are algorithmically verifiable. The methodology for computing the structural metrics Consistency and Onto-score has been improved, allowing not only the detection of node presence but also the evaluation of their coherence. Unlike previous approaches focused on template-based detection, our model performs a logical reconstruction of scientific discourse.

The practical significance of the results lies in the creation of a system capable of automatically analyzing scientific texts, detecting structural deficiencies, providing improvement recommendations, and generating preliminary reviews. The system can be integrated into editorial interfaces of scientific journals for preliminary screening of submitted articles or used as an educational tool in academic writing courses.

Among the achieved quantitative results: the precision of structural defect detection in the controlled corpus reached 82%, the Consistency score for benchmark articles was 0.92, and the Onto-score was 0.94. For articles from discontinued journals, these scores were 0.71 and 0.76, respectively, confirming the model's effectiveness in detecting both explicit and latent structural issues.

It is recommended that the obtained results be used to further automate peer review processes, structure scientific knowledge bases, and serve as an expert support tool in digital humanities and interdisciplinary research. In the future, the ontological model will be extended to other genres of scientific texts (e.g., review articles, methodological papers, dissertations), and mechanisms for explainability of logic-ontological decisions will be implemented.

REFERENCES

1. Arp R., Smith B., Spear A. D. Building ontologies with basic formal ontology. Cambridge, MIT Press, 2015, 248 p. DOI: 10.7551/mitpress/9780262527811.001.0001
2. Guizzardi G., Wagner G., Almeida J. P. A., Guizzardi R. S. S. Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story, *Applied Ontology,* 2015, Vol. 10, No. 3–4, pp. 259–271. DOI: 10.3233/AO-150157

DOI 10.15588/1607-3274-2025-4-7

3. Moltmann F. Natural Language Ontology. Oxford Research Encyclopedia of Linguistics, 2017. DOI: 10.1093/acrefore/9780199384655.013.330
4. Moltmann F. Truth predicates, truth bearers, and their variants, *Synthese,* 2021, Vol. 198, Suppl. 2, pp. 689–716. DOI: 10.1007/s11229-018-1814-8
5. Moltmann F. Relative truth and the first person, *Philosophical Studies,* 2010, Vol. 150, pp. 187–220. DOI: 10.1007/s11098-009-9383-9
6. Guizzardi G., Zamborlini V. Using a trope-based foundational ontology for bridging different areas of concern in ontology-driven conceptual modeling, *Science of Computer Programming,* 2014, Vol. 96, pp. 417–443. DOI: 10.1016/j.scico.2014.02.022
7. Cimiano P., Völker J., Studer R. Ontologies on demand? A description of the state-of-the-art, applications, challenges and trends for ontology learning from text, *Information Wissenschaft und Praxis*, 2006, Vol. 57, No. 6–7, pp. 315–320.
8. Wu S., Ma X., Luo D., Li L., Shi X., Chang X., Gong J. Automated review generation method based on large language models, *arXiv preprint*, 2024, arXiv:2407.20906. DOI: 10.48550/arXiv.2407.20906
9. Liang W., Zhang Y., Cao H., Wang B., Ding D. Y., Yang X., Zou J. Can large language models provide useful feedback on research papers? A large-scale empirical analysis, *NEJM AI,* 2024, Vol. 1, No. 8, Article AIoa2400196. DOI: 10.1056/AIoa2400196
10. Wadden D., Lin S., Lo K., Wang L. L., Zuylen M. van, Cohan A., Hajishirzi H. Fact or fiction: Verifying scientific claims, *arXiv preprint,* 2020, arXiv:2004.14974. DOI: 10.48550/arXiv.2004.14974
11. Tam D., Mascarenhas A., Zhang S., Kwan S., Bansal M., Raffel C. Evaluating the factual consistency of large language models through news summarization, *arXiv preprint*, 2022, arXiv:2211.08412. DOI: 10.48550/arXiv.2211.08412
12. Rahaman N., Weiss M., Wüthrich M., Bengio Y., Li L. E., Pal C., Schölkopf B. Language Models Can Reduce Asymmetry in Information Markets, *arXiv preprint*, 2024, arXiv:2403.14443. DOI: 10.48550/arXiv.2403.14443
13. Sinha A., Shen Z., Song Y., Ma H., Eide D., Hsu B. J., Wang K. An overview of Microsoft Academic Service (MAS) and applications, *Proceedings of the 24th International Conference on World Wide Web (WWW), 18–22 May 2015*. Florence, Italy, pp. 243–246. DOI: 10.1145/2740908.2742839
14. Elsevier. List of Scopus discontinued journals [Electronic resource]. Access mode: https://journalsearches.com/blog/scopus-discontinued-journals-list.php
15. Committee on Publication Ethics (COPE). Core practices and guidelines [Electronic resource]. Access mode: https://publicationethics.org/guidance, 2023.

Received 18.06.2025.
Accepted 22.09.2025.

УДК 004.9

ЛОГІКО-ОНТОЛОГІЧНА РЕКОНСТРУКЦІЯ ДИСКУРСИВНОЇ СТРУКТУРИ НАУКОВОГО ТЕКСТУ ТА ЇЇ РЕАЛІЗАЦІЯ В ШІ-СИСТЕМІ РЕЦЕНЗУВАННЯ

Бедратюк Л. П. – д-р фіз.-мат. наук, професор, завідувач кафедри інженерії програмного забезпечення Хмельницького національного університету, Хмельницький, Україна.

АНОТАЦІЯ

Актуальність. Зростання кількості наукових публікацій і поява інструментів на основі великих мовних моделей (LLM) актуалізують потребу в автоматизованій верифікації структурної якості наукового тексту. Більшість існуючих рішень зосереджені на поверхневому лінгвістичному аналізі та не враховують логіко-дискурсивну цілісність: зокрема, чи присутні в тексті гіпотеза, метод, результати, висновки та чи пов'язані вони між собою нормативними зв'язками.

Мета. Метою дослідження є розроблення онтологічно керованого підходу до формалізованої перевірки структури наукових текстів шляхом побудови онтологічного графа знань і оцінки його відповідності наперед визначеній нормативній моделі наукового дискурсу.

Метод. Запропоновано модель, що ґрунтується на двох взаємопов'язаних онтологіях – «Наукова публікація» (визначає типи вузлів і їх ролі) та «Рецензування» (визначає логіко-дискурсивні вимоги). Текст подається у вигляді графа у якому вузли формуються на основі семантичної розмітки за допомогою LLM, а зв'язки перевіряються відповідно до множини нормативних правил. Для реалізації використано спеціалізованого GPT-агента, здатного динамічно застосовувати онтологічні знання під час аналізу та генерації рецензій.

Результати. Модель дозволяє автоматично виявляти порушення дискурсивної структури: відсутність ключових елементів, логічну розірваність, підміну наукової новизни практичною значущістю, некоректну інтерпретацію результатів. Запропоновані метрики кількісно фіксують рівень структурної повноти та узгодженості. Наведені приклади графів і рецензій демонструють, що система здатна виявляти неочевидні, латентні порушення логіки викладу навіть у формально повних текстах.

Висновки. Наукова новизна дослідження полягає у впровадженні онтологічного графа як інтерпретованої моделі наукової аргументації, що використовується у тандемі з великою мовною моделлю. Практичне значення полягає у створенні основи для напівавтоматизованого рецензування, структурного аналізу публікацій і навчання академічному письму. Методологія є масштабованою на інші жанри наукового тексту та потенційно інтегрованою у редакційні платформи.

КЛЮЧОВІ СЛОВА: онтологія, онтологічний граф, GPT-агент, рецензування, семантичний аналіз, наукова публікація, логіко-дискурсивна структура, машинне навчання, алгоритми.

DOI 10.15588/1607-3274-2025-4-7

ЛІТЕРАТУРА

1. Arp R. Building ontologies with basic formal ontology / R. Arp, B. Smith, A. D. Spear. – Cambridge : MIT Press, 2015. – 248 p. DOI: 10.7551/mitpress/9780262527811.001.0001
2. Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story / [G. Guizzardi, G. Wagner, J. P. A. Almeida, R. S. S. Guizzardi] // Applied Ontology. – 2015. – Vol. 10, No. 3–4. – P. 259–271. DOI: 10.3233/AO-150157
3. Moltmann F. Natural Language Ontology / F. Moltmann // Oxford Research Encyclopedia of Linguistics. – 2017. DOI: 10.1093/acrefore/9780199384655.013.330
4. Moltmann F. Truth predicates, truth bearers, and their variants / F. Moltmann // Synthese. – 2021. – Vol. 198, Suppl. 2. – P. 689–716. DOI: 10.1007/s11229-018-1814-8
5. Moltmann F. Relative truth and the first person / F. Moltmann // Philosophical Studies. – 2010. – Vol. 150. – P. 187–220. DOI: 10.1007/s11098-009-9383-9
6. Guizzardi G. Using a trope-based foundational ontology for bridging different areas of concern in ontology-driven conceptual modeling / G. Guizzardi, V. Zamborlini // Science of Computer Programming. – 2014. – Vol. 96. – P. 417–443. DOI: 10.1016/j.scico.2014.02.022
7. Cimiano P. Ontologies on demand? A description of the state-of-the-art, applications, challenges and trends for ontology learning from text / P. Cimiano, J. Völker, R. Studer // Information Wissenschaft und Praxis. – 2006. – Vol. 57, No. 6–7. – P. 315–320.
8. Automated review generation method based on large language models / [S. Wu, X. Ma, D. Luo et al.] // arXiv preprint. – 2024. – arXiv:2407.20906. DOI: 10.48550/arXiv.2407.20906
9. Can large language models provide useful feedback on research papers? A large-scale empirical analysis / [W. Liang, Y. Zhang, H. Cao et al.] // NEJM AI. – 2024. – Vol. 1, No. 8. – Article AIoa2400196. DOI: 10.1056/AIoa2400196
10. Fact or fiction: Verifying scientific claims / [D. Wadden, S. Lin, K. Lo et al.] // arXiv preprint. – 2020. – arXiv:2004.14974. DOI: 10.48550/arXiv.2004.14974
11. Evaluating the factual consistency of large language models through news summarization / [D. Tam, A. Mascarenhas, S. Zhang et al.] // arXiv preprint. – 2022. – arXiv:2211.08412. DOI: 10.48550/arXiv.2211.08412
12. Language Models Can Reduce Asymmetry in Information Markets / [N. Rahaman, M. Weiss, M. Wüthrich et al.] // arXiv preprint. – 2024. – arXiv:2403.14443. DOI: 10.48550/arXiv.2403.14443
13. An overview of Microsoft Academic Service (MAS) and applications / [A. Sinha, Z. Shen, Y. Song et al.] // Proceedings of the 24th International Conference on World Wide Web (WWW), 18–22 May 2015. – Florence, Italy. – P. 243–246. DOI: 10.1145/2740908.2742839
14. Elsevier. List of Scopus discontinued journals [Electronic resource]. – Access mode: https://journalsearches.com/blog/scopus-discontinued-journals-list.php
15. Committee on Publication Ethics (COPE). Core practices and guidelines [Electronic resource]. – Access mode: https://publicationethics.org/guidance, 2023.

DOI 10.15588/1607-3274-2025-4-7