

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ

ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ  
І ГЛОБАЛЬНОГО ІНФОРМАЦІЙНОГО ПРОСТОРУ

Кваліфікаційна наукова праця  
на правах рукопису

**Гайко Світлана Іванівна**

УДК 001.89:004.5:004.91:81`3

**ДИСЕРТАЦІЯ**

**ОНТОЛОГО-КЕРОВАНІ ЗАСОБИ ПРЕДСТАВЛЕННЯ  
НЕСТРУКТУРОВАНОЇ ІНФОРМАЦІЇ**

122 – Комп'ютерні науки

Подається на здобуття кваліфікаційного рівня PhD candidate

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

\_\_\_\_\_ (С. І. Гайко)

(підпис, ініціали та прізвище здобувача)

Науковий керівник  
Стрижак Олександр Євгенійович  
доктор технічних наук, професор

Київ, 2023

## АНОТАЦІЯ

*Гайко С. І.* Онтолого-керовані засоби представлення неструктурованої інформації. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття кваліфікаційного рівня PhD candidate за спеціальністю 122 – Комп'ютерні науки. – Інститут телекомунікацій і глобального інформаційного простору Національної академії наук України, Київ, 2023.

Дисертаційне дослідження присвячено вирішенню актуальної науково-технічної задачі систематизації природномовних видань в глобальному середовищі шляхом розробки моделі, методу і засобів їх трансдисциплінарного представлення та консолідації.

**Об'єктом дослідження** є процес трансдисциплінарного представлення природномовних видань.

**Предмет дослідження** – онтологічні методи представлення видань в глобальному середовищі.

**Методи дослідження.** Проведені дослідження базуються на загальній теорії наукового пізнання та на засадах системності, множинності, рекурсивної редукції, гомотопічних типів, рефлексії текстових наративів. Для розв'язання поставлених задач було використано такі методи: для аналізу сучасних засобів представлення видань в Мережі – метод системологічної класифікації; для розробки методів розв'язання задач – теорія складних систем, системний аналіз, теорія множин, теорія графів, лямбда-числення; для формалізації представлення знань – алгебрологічний та аксіоматичний методи; для програмної реалізації інформаційної технології – шаблони проектування та об'єктно-орієнтований аналіз.

**Мета** дисертаційної роботи полягає у забезпеченні онтолого-керованого представлення неструктурованої інформації на засадах автоматичного пошуку інформації в природномовних виданнях, їх відображення у форматі онтологій та консолідації їх контекстів. Для досягнення зазначеної мети проведено наступну роботу:

**1) Розроблено концептуальну онтолого-керовану модель трансдисциплінарного представлення електронних образів природномовних видань.** Розробка даної моделі включає наступні етапи:

- Розробка таксономічної моделі структурного відображення різностильових документів;
- Розробка моделі рефлексійної взаємодії мікротекстів, як контекстів таксономічної моделі різностильових документів;
- Розробка функціональної моделі рекурсивної взаємодії між концептами таксономій різностильових документів;
- Розробка процедури перетворення таксономічних структур різностильових документів на засадах редукції;

**2) Розроблено метод трансдисциплінарного представлення природномовних видань,** основу якого становлять:

- Функціональна архітектура трансдисциплінарного образу документу;
- Концептуальна та інформаційна моделі представлення неструктурованої інформації в онтологічному форматі;
- Методика застосування спеціалізованого додатку КІТ «Поліедр» – Редуктор;
- Класифікація різностильових документів;
- Шкала ступеня формалізованості документу;
- Модель трансдисциплінарної консолідації інформаційних ресурсів;
- Процедура формування онтологічного опису системи документів на основі його текстового представлення;
- Правила автоматизованого виділення інформації з документів.

**3) Створено мережевий програмно-інформаційні засіб трансдисциплінарного представлення інформації, що міститься в електронних виданнях.** Створений засіб дозволяє:

- відбирати та накопичувати інформацію з текстів природною мовою;
- будувати тематичні глосарії, тезауруси, тощо;
- класифікувати та каталогізувати тематичну інформацію;

– здійснювати онтологічне управління та дослідницький інжиніринг, які забезпечують встановлення ієрархічних структур на всіх рівнях мовно-онтологічного опису тематики предметних галузей;

– будувати категоріальні рівні мовно-онтологічних описів тематик предметних сфер на основі використання засобів онтологічного моделювання;

– консолідувати контексти різностильових документів у процесі їх використання користувачем;

– вилучати із множини текстових документів знання, релевантні до заданої предметної галузі, здійснювати їх системно-онтологічної структуризації та формально-логічного подання.

В процесі виконання дисертаційної роботи отримано такі **основні результати**:

– Проаналізовано сучасні засоби представлення природномовних видань в Мережі, розроблено їх класифікацію.

– Розроблено метод трансдисциплінарного представлення природномовних видань;

– На основі онтолого-керованого підходу удосконалено технологічні аспекти об'єктно-орієнтованого підходу до пошуку інформації;

– Розроблено ІТ-технологію формування онтологій видань на основі репрезентації їх таксономій, зв'язків між їх контекстами та множинного представлення видань, що забезпечує ефективний пошук інформації у природномовних виданнях;

– Створено онтологічний репозиторій нормативних документів;

– Набули подальшого розвитку ідеї ефективною структуризації та, відповідно, швидкого пошуку інформації.

**Ключові слова:** онтологія, онтологічне моделювання, трансдисциплінарність, природномовні видання, різностильові документи, контекст, інтерактивний документ, знання-орієнтована система.

*Список публікацій здобувача*

*Публікації у наукових фахових виданнях України*

1. Стрижак, О.Є., Приходнюк В.В., **Гайко С. І.**, Шаповалов В.Б. Відображення мережевої інформації у вигляді інтерактивних документів. Трансдисциплінарний підхід. *Математичне моделювання в економіці*. 2018. № 3 (12). С. 87-100.
2. Приходнюк В.В., Стрижак О.Є., **Гайко С. І.**, Чепков Р.І. Інформаційно-аналітичний комплекс підтримки процесів трансдисциплінарних досліджень. *Екологічна безпека та природокористування*. 2018. № 4 (28). С. 103-119.
3. **Гайко С. І.**, Приходнюк В. В. Підхід до автоматизованої структуризації освітніх ресурсів на основі методу рекурсивної редукції. *Наукові записки Малої академії наук України*. 2021. № 1 (19). С. 28-38. DOI: <http://doi.org/10.51707/2618-0529-2021-20-03>.
4. **Гайко С. І.** Онтолого-керовані засоби оброблення та подання великих масивів неструктурованих текстів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 2 (24). С. 27-38. DOI: <https://doi.org/10.30837/ITSSI.2023.24.027>.

*Публікації у виданнях іноземних держав, що входять до організації економічного співробітництва та розвитку Європейського союзу*

5. **Гайко С. І.**, Приходнюк В. В. Средства трансдисциплінарного представлення інформаційних ресурсів різних стилей. *Information Models and Analyses*. 2020. Vol. 9, No 1. P. 78-99.

*Статті у наукових закордонних виданнях, що індексуються у міжнародних наукометричних базах (Scopus, Web of Science)*

6. Stryzhak O., Prychodniuk V., Popova M., Nadutenko M., **Haiko S.**, Chepkov R. (2021) Development of an Oceanographic Databank Based on Ontological Interactive Documents. In: Arai K. (eds) *Intelligent Computing. Lecture Notes in Networks and Systems*. Vol. 284. Springer, Cham. P. 97-114. DOI: [https://doi.org/10.1007/978-3-030-80126-7\\_8](https://doi.org/10.1007/978-3-030-80126-7_8).

Матеріали та тези конференцій

7. **Гайко С. І.**, Приходнюк В. В. Автоматизована обробка академічних видань засобами системи «Рекурсивний редуктор». *Інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях*: колективна монографія за мат. XVII міжнар. наук.-практ. конф. (м. Київ, 25-26 вересня 2018). Київ, 2018. С. 126-128.
8. Гуржій А., **Гайко С.**, Попова М., Стрижак О. Трансдисциплінарні когнітивні засоби підтримки наукових досліджень. *Освіта для миру = Edukacja dla pokoju*: зб. наук. пр. VIII українсько-польського наук. форуму (м. Київ, 8-10 жовтня 2019 р.). Київ, 2019. Т.1. С. 190-200.
9. **Гайко С. І.**, Приходнюк В. В. Обробка інформації в середовищі «ТОДОС-Редуктор» на прикладі тексту публіцистичного стилю. *Радіоелектроніка та молодь у XXI столітті*: матеріали 23-го міжнар. молодіжного форуму (м. Харків, 16-18 квітня 2019 р.). Харків, 2019. С.11-12.
10. **Гайко С. І.** Методика застосування системи «Рекурсивний редуктор». *Сучасні технології в науці та освіті*: матеріали другої міжнар. науково-практ. конф. (м. Северодонецьк, 5–7 березня 2019 р.). Северодонецьк, 2019. С. 90-92.
11. **Гайко С. І.**, Приходнюк В. В. Редуктивні правила обробки відомчої інформації. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях*: колективна монографія за матеріалами XVIII міжнар. наук.-практ. конф. (м. Київ, 01-02 жовтня 2019 р.). Київ, 2019. С. 196-199.
12. **Гайко С. І.**, Приходнюк В. В. Застосування методу рекурсивної редукції для обробки великих масивів різнорідних документів. *Радіоелектроніка та молодь у XXI столітті*: матеріали 24-го Міжнар. молодіжного форуму (м. Харків, 7-9 квітня 2020 р.). Харків, 2020. С. 181-182.
13. **Гайко С. І.**, Приходнюк В. В. Представлення освітніх ресурсів у вигляді інтерактивних документів. *Інтернет – Освіта – Наука – 2020*: зб. праць XII міжнар. наук.-практ. конф. (м. Вінниця, 26 – 29 травня 2020 р.). Вінниця, 2020. С. 249-251.
14. **Гайко С. І.**, Приходнюк В. В. Засоби трансдисциплінарного представлення освітніх ресурсів. *Актуальні аспекти розвитку STEM-освіти у навчанні*

*природничо-наукових дисциплін: зб. мат. III міжнар. наук.-практ. конф. (м. Кропивницький, 14-15 травня 2020 р.). Кропивницький, 2020. С. 61-64.*

15. **Гайко С. І.**, Зотова Л. В. Структуризація інформаційних ресурсів різних стилів у форматі наративного дискурсу. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: тенденції 2020 року: колективна монографія за матеріалами XIX міжнар. наук.-практ. конф. (м. Київ, 06-07 жовтня 2020 р.). Київ, 2020. С. 175-177.*
16. **Гайко С. І.**, Стрижак О. Є. Модель трансдисциплінарної інтеграції інформаційних ресурсів різних стилів. *Ідеї академіка В.М. Глушкова і сучасні проблеми теоретичної кібернетики: матеріали IX всеукр. наук.-практ. конф. (м. Київ, 18 грудня 2020 р.). Київ, 2020. С. 23-26.*
17. **Гайко С. І.** Застосування системи автоматизованої обробки різностильових документів на прикладі океанографічних законодавчих актів. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: виклики 2021 року: колективна монографія за матеріалами 20 міжнар. наук.-практ. конф. (м. Київ, 04-08 жовтня 2021 р.). Київ, 2021. С. 203-206.*

Авторське свідоцтво

18. А. с. Комп'ютерна програма «Трансдисциплінарні онтологічні дослідження операціональних середовищ і процесів» («ТОДОС-Процеси») / О. Є. Стрижак, В. Ю. Величко, В. В. Приходнюк, М. В. Надутенко, В. В. Горборуков, О. В. Франчук, С. О. Довгий, О. В. Лісовий, О. В. Палагін, І. В. Сергієнко, В. А. Широков, Ю. О. Романенков, І. В. Шостак, І. Б. Чепков, А. Ю. Гупало, Г. М. Потапов, Д. А. Філістеєв, Л. С. Глоба, О. В. Гладун, Т. О. Грязнухіна, М. Ю. Кригін, Т. П. Любченко, І. В. Шевченко, Л. А. Шипіцина, **С. І. Гайко**, К. С. Малахов (Україна). - № 96130; Дата реєстрації 18.02.2020.

## ABSTRACT

*Haiko S. I.* Ontology-driven means of presenting unstructured information – Qualifying scientific work on manuscript rights.

Thesis for the PhD Candidate degree on the specialty 122 - «Computer Science». Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, Kyiv, 2023.

The thesis is dedicated to solving the current scientific and technical task of systematization of natural language publications in a global environment by developing a model, method and means of their transdisciplinary presentation and consolidation.

**The object** of research is the process of transdisciplinary presentation of natural language publications.

**The subject** of research is ontological methods of presenting publications in a global environment.

**Research methods.** The conducted research is based on the general theory of scientific knowledge and on the principles of systematicity, multiplicity, recursive reduction, homotopy types, reflection of textual narratives. The following methods were used to solve the problems: for the analysis of modern means of presenting publications on the Internet - the method of systemological classification; for the development of problem solving methods - theory of complex systems, system analysis, set theory, graph theory, lambda calculus; for the formalization of knowledge presentation – algebraic-logical and axiomatic methods; for software implementation of information technology – design templates and object-oriented analysis.

**The purpose** of the thesis is to provide an ontology-driven presentation of unstructured information based on the automatic search for information in natural language publications, their display in the format of ontologies, and the consolidation of their contexts. To achieve the specified goal, the following work was carried out:

**1) A conceptual ontology-driven model of transdisciplinary representation of electronic images of natural language publications has been developed.** The development of this model includes the following stages:



- Development of a taxonomic model of the structural display of documents of different styles;
- Development of a model of reflexive interaction of microtexts as contexts of a taxonomic model of documents of various styles;
- Development of a functional model of recursive interaction between concepts of taxonomies of different styles of documents;
- Development of a procedure for transforming taxonomic structures of various style documents on the basis of reduction;

**2) A method of transdisciplinary presentation of natural language publications has been developed,** the basis of which is:

- Functional architecture of the transdisciplinary image of the document;
- Conceptual and information models of unstructured information presentation in ontological format;
- Application method of the specialized application KIT "Polyedr" – Reducer;
- Classification of various style documents;
- The scale of the degree of formalization of the document;
- Model of transdisciplinary consolidation of information resources;
- The procedure for forming an ontological description of the document system based on its text representation;
- Rules for automated extraction of information from documents.

**3) A network software and information tool for transdisciplinary presentation of information contained in electronic publications has been created.** The created tool allows:

- select and accumulate information from texts in natural language;
- build thematic glossaries, thesauri, etc.;
- classify and catalog thematic information;
- carry out ontological management and research engineering, which ensure the establishment of hierarchical structures at all levels of linguistic and ontological description of subject areas;

- build categorical levels of linguistic and ontological descriptions of subject areas based on the use of ontological modeling tools;
- consolidate the contexts of documents of different styles in the process of their use by the user;
- extract knowledge relevant to a given subject area from a set of text documents, carry out their system-ontological structuring and formal-logical presentation.

In the course of the dissertation work, the following **main results** were obtained:

- Modern means of presenting natural language publications on the Web were analyzed, and their classification was developed;
- A method of transdisciplinary presentation of natural language publications was developed;
- The technological aspects of the object-oriented approach to information retrieval have been improved;
- The IT technology for the formation of ontologies of publications based on the representation of their taxonomies, connections between their contexts and multiple representation of publications was developed, which ensures an effective search for information in natural language publications;
- An ontological repository of normative documents has been created;
- The idea of effective structuring and, accordingly, quick search for information gained further development.

**Keywords:** ontology, ontological modeling, transdisciplinarity, natural language editions, multi-style documents, context, interactive document, knowledge-oriented system.

*List of the candidate's publications*

1. Стрижак, О.Є., Приходнюк В.В., **Гайко С. І.**, Шаповалов В.Б. Відображення мережевої інформації у вигляді інтерактивних документів. Трансдисциплінарний підхід. *Математичне моделювання в економіці*. 2018. № 3 (12). С. 87-100.
2. Приходнюк В.В., Стрижак О.Є., **Гайко С. І.**, Чепков Р.І. Інформаційно-аналітичний комплекс підтримки процесів трансдисциплінарних досліджень. *Екологічна безпека та природокористування*. 2018. № 4 (28). С. 103-119.
3. **Гайко С. І.**, Приходнюк В. В. Підхід до автоматизованої структуризації освітніх ресурсів на основі методу рекурсивної редукції. *Наукові записки Малої академії наук України*. 2021. № 1 (19). С. 28-38. DOI: <http://doi.org/10.51707/2618-0529-2021-20-03>.
4. **Гайко С. І.** Онтолого-керовані засоби оброблення та подання великих масивів неструктурованих текстів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 2 (24). С. 27-38. DOI: <https://doi.org/10.30837/ITSSI.2023.24.027>.
5. **Гайко С. І.**, Приходнюк В. В. Средства трансдисциплинарного представления информационных ресурсов разных стилей. *Information Models and Analyses*. 2020. Vol. 9, No 1. P. 78-99.
6. Stryzhak O., Prychodniuk V., Popova M., Nadutenko M., **Haiko S.**, Chepkov R. (2021) Development of an Oceanographic Databank Based on Ontological Interactive Documents. In: Arai K. (eds) *Intelligent Computing. Lecture Notes in Networks and Systems*. Vol. 284. Springer, Cham. P. 97-114. DOI: [https://doi.org/10.1007/978-3-030-80126-7\\_8](https://doi.org/10.1007/978-3-030-80126-7_8). ISSN 2367-3389 (Online)
7. **Гайко С. І.**, Приходнюк В. В. Автоматизована обробка академічних видань засобами системи «Рекурсивний редуктор». *Інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: колективна монографія за мат. XVII міжнар. наук.-практ. конф. (м. Київ, 25-26 вересня 2018)*. Київ, 2018. С. 126-128.
8. Гуржій А., **Гайко С.**, Попова М., Стрижак О. Трансдисциплінарні когнітивні засоби підтримки наукових досліджень. *Освіта для миру = Edukacja dla pokoju: зб. наук. пр. VIII українсько-польського наук. форуму (м. Київ, 8-10 жовтня 2019 р.)*. Київ, 2019. Т.1. С. 190-200.

9. **Гайко С. І.,** Приходнюк В. В. Обробка інформації в середовищі «ТОДОС-Редуктор» на прикладі тексту публіцистичного стилю. *Радіоелектроніка та молодь у XXI столітті: матеріали 23-го міжнар. молодіжного форуму* (м. Харків, 16-18 квітня 2019 р.). Харків, 2019. С. 11-12.
10. **Гайко С. І.** Методика застосування системи «Рекурсивний редуктор». *Сучасні технології в науці та освіті: матеріали другої міжнар. науково-практ. конф.* (м. Сєверодонецьк, 5–7 березня 2019 р.). Сєверодонецьк, 2019. С. 90-92.
11. **Гайко С. І.,** Приходнюк В. В. Редуктивні правила обробки відомчої інформації. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: колективна монографія за матеріалами XVIII міжнар. наук.-практ. конф.* (м. Київ, 01-02 жовтня 2019 р.). Київ, 2019. С. 196-199.
12. **Гайко С. І.,** Приходнюк В. В. Застосування методу рекурсивної редукції для обробки великих масивів різнорідних документів. *Радіоелектроніка та молодь у XXI столітті: матеріали 24-го Міжнар. молодіжного форуму* (м. Харків, 7-9 квітня 2020 р.). Харків, 2020. С. 181-182.
13. **Гайко С. І.,** Приходнюк В. В. Представлення освітніх ресурсів у вигляді інтерактивних документів. *Інтернет – Освіта – Наука – 2020: зб. праць XII міжнар. наук.-практ. конф.* (м. Вінниця, 26 – 29 травня 2020 р.). Вінниця, 2020. С. 249-251.
14. **Гайко С. І.,** Приходнюк В. В. Засоби трансдисциплінарного представлення освітніх ресурсів. *Актуальні аспекти розвитку STEM-освіти у навчанні природничо-наукових дисциплін: зб. мат. III міжнар. наук.-практ. конф.* (м. Кропивницький, 14-15 травня 2020 р.). Кропивницький, 2020. С. 61-64.
15. **Гайко С. І.,** Зотова Л. В. Структуризація інформаційних ресурсів різних стилів у форматі наративного дискурсу. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: тенденції 2020 року: колективна монографія за матеріалами XIX міжнар. наук.-практ. конф.* (м. Київ, 06-07 жовтня 2020 р.). Київ, 2020. С. 175-177.
16. **Гайко С. І.,** Стрижак О. Є. Модель трансдисциплінарної інтеграції інформаційних ресурсів різних стилів. *Ідеї академіка В.М. Глушкова і сучасні*

*проблеми теоретичної кібернетики: матеріали ІХ всеукр. наук.-практ. конф. (м. Київ, 18 грудня 2020 р.). Київ, 2020. С. 23-26.*

17. **Гайко С. І.** Застосування системи автоматизованої обробки різностильових документів на прикладі океанографічних законодавчих актів. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: виклики 2021 року: колективна монографія за матеріалами 20 міжнар. наук.-практ. конф. (м. Київ, 04-08 жовтня 2021 р.). Київ, 2021. С. 203-206.*
18. А. с. Комп'ютерна програма «Трансдисциплінарні онтологічні дослідження операціональних середовищ і процесів» («ТОДОС-Процеси») / О. Є. Стрижак, В. Ю. Величко, В. В. Приходнюк, М. В. Надутенко, В. В. Горборуков, О. В. Франчук, С. О. Довгий, О. В. Лісовий, О. В. Палагін, І. В. Сергієнко, В. А. Широков, Ю. О. Романенков, І. В. Шостак, І. Б. Чепков, А. Ю. Гупало, Г. М. Потапов, Д. А. Філістєєв, Л. С. Глоба, О. В. Гладун, Т. О. Грязнухіна, М. Ю. Кригін, Т. П. Любченко, І. В. Шевченко, Л. А. Шипіцина, **С. І. Гайко**, К. С. Малахов (Україна). - № 96130; Дата реєстрації 18.02.2020.

## ЗМІСТ

<b>ПЕРЕЛІК ОСНОВНИХ СКОРОЧЕНЬ.....</b>	<b>17</b>
<b>ВСТУП .....</b>	<b>18</b>
<b>РОЗДІЛ 1. АНАЛІЗ ІСНУЮЧИХ СИСТЕМ ОБРОБКИ ТА ПРЕДСТАВЛЕННЯ ПРИРОДНОМОВНИХ ТЕКСТІВ.....</b>	<b>25</b>
1.1. Огляд моделей систем автоматизованої обробки та представлення природномовних текстів .....	25
1.1.1. <i>Моделі систем обробки природномовних текстів.....</i>	<i>25</i>
1.1.2. <i>Моделі представлення природномовної інформації .....</i>	<i>34</i>
1.2. Трансдисциплінарність як критерій формування сучасних систем обробки та представлення природномовних текстів.....	39
1.3. Порівняння існуючих систем обробки та представлення природномовних текстів .....	43
1.4. Висновки за розділом 1.....	50
<b>РОЗДІЛ 2. МОДЕЛЬ ТРАНСДИСЦИПЛІНАРНОГО ПРЕДСТАВЛЕННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ РІЗНИХ СТИЛІВ.....</b>	<b>51</b>
2.1. Загальні поняття .....	51
2.1.1. <i>Онтології.....</i>	<i>51</i>
2.1.2. <i>Таксономічні основи структуризації документів .....</i>	<i>55</i>
2.1.3. <i>Зростаюча пірамідальна мережа .....</i>	<i>61</i>
2.1.4. <i>Метод рекурсивної редукції .....</i>	<i>65</i>
2.1.5. <i>Онтологічний інтерактивний документ.....</i>	<i>69</i>
2.2. Концептуальна онтолого-керована модель трансдисциплінарного представлення ІР.....	70
2.2.1. <i>Модель онтолого-керованої класифікації та ідентифікації інформації.....</i>	<i>71</i>
2.2.2. <i>Модель класифікації природномовного тексту за стилем.....</i>	<i>74</i>
2.2.3. <i>Модель трансдисциплінарної консолідації інформаційних ресурсів .....</i>	<i>76</i>

2.3.	Формалізація документів різних стилів на основі рекурсивної редуції.....	83
2.3.1.	<i>Класифікація функціональних стилів.....</i>	83
2.3.2.	<i>Шкала ступеню формалізованості документу.....</i>	87
2.3.3.	<i>Оцінка еквівалентності документів різних стилів з точки зору автоматизованої обробки .....</i>	94
2.4.	Висновки за розділом 2 .....	95

## **РОЗДІЛ 3. АРХІТЕКТУРА ПРОГРАМНОЇ СИСТЕМИ**

### **ТРАНСДИСЦИПЛІНАРНОГО ПРЕДСТАВЛЕННЯ**

### **ІНФОРМАЦІЙНИХ РЕСУРСІВ РІЗНИХ СТИЛІВ..... 96**

3.1.	Інформаційна модель програмної системи трансдисциплінарного представлення різностильових ІР .....	97
3.2.	Модель поведінки системи .....	101
3.2.1.	<i>Варіанти використання системи .....</i>	101
3.2.2.	<i>Активність системи .....</i>	103
3.2.3.	<i>Взаємодія системи .....</i>	105
3.3.	Архітектура системи .....	106
3.4.	Програмний засіб КІТ «Поліедр». Загальна структура .....	108
3.5.	Висновки за розділом 3 .....	110

## **РОЗДІЛ 4. ПРОГРАМНЕ СЕРЕДОВИЩЕ ТРАНСДИСЦИПЛІНАРНОГО**

### **ПРЕДСТАВЛЕННЯ ПРИРОДНОМОВНИХ ВИДАНЬ .....112**

4.1.	Онтологія стилів та її формування за допомогою стандартних засобів КІТ «Поліедр» .....	112
4.1.1.	<i>Функція створення об'єкту.....</i>	113
4.1.2.	<i>Функція створення зв'язків .....</i>	114
4.1.3.	<i>Функція видалення об'єктів і зв'язків .....</i>	115
4.1.4.	<i>Функція редагування атрибутів.....</i>	117
4.1.5.	<i>Функція зчитування атрибутів з XLSX-файлів .....</i>	122
4.1.6.	<i>Функція збереження в бібліотеку .....</i>	123

4.1.7.	<i>Функція роботи з буфером обміну</i> .....	127
4.2.	Онтологічний шаблон представлення для подальшої обробки результатів аналізу .....	129
4.3.	Модуль класифікації та обробки інформації .....	133
4.3.1.	<i>Загальна структура модулю класифікації та обробки інформації</i> .....	133
4.4.	Трансдисциплінарна інтеграція інформації .....	136
4.4.1.	<i>Робота з індексними зонами в середовищі КІТ «Поліедр»</i> .....	136
4.4.2.	<i>Пошук та представлення інформації в середовищі КІТ «Поліедр»</i> .....	138
4.5.	Приклад трансдисциплінарного представлення великого масиву неструктурованих текстів .....	144
4.6.	Приклад класифікації та вибору контенту .....	149
4.7.	Приклад формування критеріїв вибору інформації .....	154
4.8.	Оцінка ефективності впровадження створених онтолого-керуваних засобів представлення неструктурованої інформації .....	158
4.9.	Висновки за розділом 4 .....	160
	<b>ВИСНОВКИ</b> .....	<b>162</b>
	<b>СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ</b> .....	<b>163</b>
	<b>Додаток А. Акти впроваджень</b> .....	<b>178</b>
	<b>Додаток Б. Список опублікованих праць за темою дисертації</b> .....	<b>181</b>
	<b>Додаток В. Відомості про апробацію результатів дисертації</b> .....	<b>185</b>
	<b>Додаток Г. Експертна оцінка підстилів</b> .....	<b>186</b>



**ПЕРЕЛІК ОСНОВНИХ СКОРОЧЕНЬ**

ІАС	Інформаційно-аналітична система
ПМТ	Природномовний текст
ІР	Інформаційні ресурси
ЗПМ	Зростаюча пірамідальна мережа
ПГ	Предметна галузь
КІТ	Когнітивна інформаційна технологія
ОКЗПНІ	Онтолого-керовані засоби представлення неструктурованої інформації
СППР	Система підтримки прийняття рішень

## ВСТУП

**Обґрунтування вибору теми дослідження.** Сучасний інформаційний простір характеризується стрімким зростанням об'ємів інформації різного формату, змісту та тематичних профілів. Цей процес породжує ряд проблем, які пов'язані з конструктивним використанням необхідної інформації для експертів у ході реалізації різноманітних проектів, що носять комплексний характер. Сюди слід віднести пошук, обробку, класифікацію, категоризацію, структурування, інтеграцію та консолідацію ресурсів глобального інформаційного середовища. Також слід підкреслити, що проблематика інтеграції та консолідації інформаційних ресурсів носить широкий і розмитий характер завдяки гетерогенності (неоднорідності), розподіленості та неповноті.

Розв'язання цих проблем лежить у напрямках, пов'язаних зі створенням та використанням різноманітних засобів обробки інформації, як пасивної системи мережових знань.

Зазначимо, що на сьогодні спостерігається розрив між достатньо розвиненими засобами граматичної обробки природномовної інформації для вирішення вузькоспеціалізованих задач і недостатністю таких засобів для вирішення комплексних задач пов'язаних з функціональним аналізом і розумінням ПМ-інформації, її формально-логічним представленням, добуванням предметних знань з їх наступним використанням.

Саме тому найбільш перспективними напрямками створення та використання систем обробки та представлення інформації вважаються ті, які здатні обробляти розподілені, політематичні, великі масиви слабо або неструктурованих даних (зокрема, природномовних документів), і тим самим надавати технологічну підтримку спеціалістам різних галузей в ефективному та конструктивному застосуванні накопичених людством знань.

При проведенні дисертаційного дослідження автор спирався на фундаментальні роботи в області складних та знання-орієнтованих інформаційних систем, що розглядалися у дослідженнях Глушкова В. М., Гладуна В. П.,

Палагіна О. В., Широкова В. А., Хорошевського В. Ф., Поспелова Д. А., Овдій О. М., Гаврилової Т. А., Загорулька Ю. А., Малишевського А. В., Андона П. І., Валькмана Ю. Р., Соловійової К. О., Стрижака О. Є., Яловця А. Л., Gruber T., Guarino N., Noy N., Corcho O., Gomez-Perez A., Gruninger M., Fernandez-Lopez M., Happel H., McGuinness D., Lee J. та ін.; а також засобів трансдисциплінарної взаємодії Князевої О. М., Лисака Л. В., Мокія М. С., Палагіна О. В., Пригожина І. Р., Стрижака О. Є., Широкова В. А., Haken H., Kahneman D., Köhler W., Morin E., Nicolescu B., Piage J., Pink D., Polak F., Rose J., Stengers I., Whole A. та ін.

Таким чином, дисертаційне дослідження має важливе наукове та практичне значення, оскільки спрямоване на вирішення актуальної науково-технічної задачі систематизації різнорідних природномовних видань в глобальному середовищі шляхом розробки моделі, методу і засобів їх трансдисциплінарного представлення та консолідації.

**Зв'язок з науковими програмами.** Дисертацію виконано в межах наукових тем Інституту телекомунікацій і глобального інформаційного простору: «Створення програмно-інформаційних засобів інформаційно-аналітичного забезпечення мережецентричних ситуаційних центрів», (ДР № 0116U000794), 2016–2020 рр.; «Створення методів та технологічних засад формування інтерактивних баз знань», (ДР№017U000005), 2017–2021 рр.; “Розробка інтерактивної бази знань наукових здобутків установ НАН України для забезпечення навчально-дослідницької діяльності учнівської молоді” (0120U100083), 2020-2022 рр.

**Мета і завдання дослідження.** Мета дисертаційної роботи полягає у забезпеченні онтолого-керованого представлення неструктурованої інформації на засадах автоматичного пошуку інформації в природномовних виданнях, їх відображення у форматі онтологій та консолідації їх контекстів.

Відповідно до мети роботи поставлені такі завдання дослідження:

– проаналізувати сучасні засоби представлення різних видів видань в Мережі, розробити їх класифікацію та охарактеризувати їх;

- розробити концептуальну онтологічну модель і засоби трансдисциплінарного представлення електронних видань;
- розробити методи трансдисциплінарного представлення різностильових природномовних видань;
- створити мережеві програмно-інформаційні засоби трансдисциплінарного представлення інформації, що міститься в електронних виданнях;
- експериментально підтвердити ефективність роботи розроблених моделей і програмно-інформаційних засобів.

**Об’єктом досліджень** є процес трансдисциплінарного представлення природномовних видань.

**Предмет досліджень** – онтологічні методи представлення видань в глобальному середовищі.

**Методи дослідження.** Проведені дослідження базуються на загальній теорії наукового пізнання та на засадах системності, множинності, рекурсивної редукції, гомотопічних типів, рефлексії текстових наративів. Для розв’язання поставлених задач було використано такі методи: для аналізу сучасних засобів представлення книг в Мережі – метод системологічної класифікації; для розробки методів розв’язання задач – теорія складних систем, системний аналіз, теорія множин, теорія графів, лямбда-числення; для формалізації представлення знань – алгебрологічний та аксіоматичний методи; для програмної реалізації інформаційної технології – шаблони проектування та об’єктно-орієнтований аналіз.

**Наукова новизна та теоретичне значення одержаних результатів:**

- Набула подальшого розвитку теорія трансдисциплінарності інформації. Розроблено ряд моделей її онтологічного представлення, оцінювання, та трансдисциплінарної консолідації. Визначено рекурсивні та редуктивні властивості різностильових документів;
- Вперше визначено модель та метод трансдисциплінарної консолідації різностильових документів;

– Вперше розроблено онтологічний класифікатор стилів, використання якого забезпечує підвищення ефективності семантичного оброблення наративів та документів різних тематичних профілів, а також реалізує виявлення їх міжконтекстної зв'язності на основі формування взаємозв'язаних онтологій, які відображають їх структуру;

– Вперше створено модель онтолого-керованої ідентифікації інформації, яка за рахунок онтології стилів забезпечила підвищення ефективності процесу створення бази правил редукції;

– Розроблена ІТ-технологія формування онтологій видань на основі репрезентації їх таксономій, зв'язків між їх контекстами та множинного представлення документів, що забезпечує ефективний пошук інформації у природномовних джерелах;

– Розроблено шкалу формалізованості різностильових документів на основі рекурсивної редукції;

– На основі онтології задачі вибору розроблено спосіб оцінки сучасних засобів представлення електронних образів видань, що дозволило виявити переваги Рекурсивного Редуктора для автоматизованої обробки природномовних текстів.

#### **Практичне значення дисертаційного дослідження:**

– розроблено програмний засіб трансдисциплінарного представлення різномірної інформації, що дозволяє здійснювати витяг із множини наративів та документів, що відображають певні знання, релевантних до заданої предметної області, їх системно-онтологічної структуризації та формально-логічного подання;

– розвинуто технологічні аспекти об'єктно-орієнтованого підходу до пошуку інформації;

–забезпечено можливості більш швидкого пошуку інформації завдяки утворенню систем онтологій;

– створено репозиторій електронного представлення законодавчих актів.

**Практичне значення одержаних результатів дослідження підтверджується** актами впроваджень (Додаток А).

**Особистий внесок здобувача** полягає в наступному:

Всі результати, які становлять основний зміст дисертації, автор отримав самостійно. Роботи [4, 10, 17] зі списку публікацій здобувача (Додаток Б) виконувалися без співавторів. В наукових працях, опублікованих у співавторстві, автору дисертації належить: таксономічне відображення структури мережевих документів та їх контекстної зв'язності між усіма концептами-поняттями, які формують термінополе користувача, на засадах рекурсивної редукції [1], механізми формування єдиного інформаційного середовища на основі WEB-орієнтованих сервісів ІАС з використанням множинних характеристик таксономій предметних галузей, що аналізуються [2], аналіз найбільш поширених програмних засобів, що призначені для обробки природномовних текстів, спосіб оцінки даних програмних засобів на основі багатокритеріальної оптимізації, автоматизована структуризація освітнього інформаційного ресурсу, представленого навчальною програмою з хімії (поглиблене навчання), відображення її у вигляді онтології, а в подальшому – створення на її основі трансдисциплінарного інтерактивного документу засобами КІТ «Поліедр» [3], опис ІАС, призначених для обробки та відображення різних інформаційних ресурсів, механізм обробки текстів природною мовою, що передуює створенню таких систем [5], підхід до розвитку океанографічного банку даних на основі формування онтологічного представлення інформаційних масивів та доступу до них за допомогою онтологічно визначеного інтерфейсу [6], технологія автоматизованої обробки ПМТ різних стилів і формування на їх основі інтерактивних документів, механізм відображення їх структури, яким є нарративний дискурс [7], онтологічний підхід до консолідації та агрегації автоматизованих ІАС з використанням найбільш поширених технологій і стандартів проектування і реалізації знання-орієнтованих розподілених систем [8], опис рекурсивної редукції як технологічної основи для створення онтологій, обробка статті публіцистичного стилю [9], процедура створення онтологій за допомогою рекурсивного редуктора [11], вдосконалення методу рекурсивної редукції шляхом створення онтологічного класифікатора стилів документів [12],

механізм автоматизованої структуризації освітніх ресурсів, для відображення їх в інтерактивній формі [13], обґрунтування трансдисциплінарного представлення освітніх ресурсів [14], модель структуризації інформаційних ресурсів різних стилів у форматі наративного дискурсу [15], модель поведінки системи (у вигляді UML-діаграм), а також архітектура програмного засобу трансдисциплінарної консолідації інформаційних ресурсів різних стилів [16], випробування системи автоматизованої обробки різностильових документів на прикладі океанографічних законодавчих актів [18].

**Апробація результатів дисертації.** Основні наукові положення та результати досліджень, викладені в дисертаційній роботі, доповідалися і обговорювалися на 11 науково-практичних конференціях, а саме на: XVII Міжнар. наук.-практ. конф. Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях (Київ, 25-26 вересня 2018); 23 Міжнародний молодіжний форум Радіоелектроніка та молодь у XXI столітті (Харків, 16-18 квітня 2019 р.); II Міжнар. наук.-практ. конф. Сучасні технології в науці та освіті (Сєверодонецьк, 5–7 березня 2019 р.); XVIII Міжнар. наук.-практ. конф. Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях (Київ, 01-02 жовтня 2019); XXIV Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті» (Харків, 15-16 жовтня 2020 р.); XII Міжнар. наук.-практ. конф. «Інтернет – Освіта – Наука – 2020» (м. Вінниця, 26 – 29 травня 2020 р.); III Міжнар. наук.-практ. конф. «Актуальні аспекти розвитку STEM-освіти у навчанні природничо-наукових дисциплін» (м. Кропивницький, 14-15 травня 2020 р.); XIX Міжнар. наук.-практ. конференції Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях (Київ, 06-07 жовтня 2020 р.); IX Всеукр. наук.-практ. конф. «Глушковські читання» (Київ, 18 грудня 2020 р.); XX Міжнар. наук.-практ. конф. Сучасні інформаційні технології управління екологічною безпекою,

природокористуванням, заходами в надзвичайних ситуаціях (Київ, 04-08 жовтня 2021). Інформацію про форму участі в конференціях наведено у Додатку В.

**Публікації.** За темою дисертації з викладенням її основних результатів опубліковано 18 наукових праць, серед яких 4 – у наукових фахових виданнях України (в тому числі 1 – одноосібно), 1 – в науковому періодичному виданні іноземних держав, що входять до Європейського союзу, 1 наукова праця у періодичному виданні, що входить до наукометричної бази SCOPUS, 11 – в матеріалах і тезах міжнародних та всеукраїнських конференцій (в тому числі 2 – одноосібно), а також 1 авторське свідоцтво. Із праць, що опубліковано у співавторстві, в дисертаційній роботі використано виключно ті результати, які одержано здобувачем особисто.

**Обсяг і структура дисертації.** Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних літературних джерел зі 154 найменувань та 4 додатків. Повний обсяг дисертації складає 187 сторінок машинописного тексту. Основний зміст викладений на 163 сторінках. Робота ілюстрована 65 рисунками та 10 таблицями.



## РОЗДІЛ 1. АНАЛІЗ ІСНУЮЧИХ СИСТЕМ ОБРОБКИ ТА ПРЕДСТАВЛЕННЯ ПРИРОДНОМОВНИХ ТЕКСТІВ

### 1.1. Огляд моделей систем автоматизованої обробки та представлення природномовних текстів

#### 1.1.1. Моделі систем обробки природномовних текстів

Природномовні тексти і досі лишаються найпоширенішою формою представлення знань. Переважна більшість видань, від законодавчих актів до художніх творів, створюються людьми на природній мові і у такому, неструктурованому вигляді зберігається у величезній кількості в глобальному цифровому середовищі. Проте, проаналізувати подібні інформаційні масиви і реалізувати їх цільове використання людям, без машинних засобів, не під силу.

У зв'язку з цим виникає необхідність в розробці систем автоматизованого аналізу масивів природномовних текстів для рішення поставлених задач. Коло таких задач доволі широке. Серед них виділяють: автоматичний машинний переклад з однієї людської мови на іншу; інформаційний пошук; видобування даних; системи, спроможні відповідати на запитання користувачів, які вони формулюють лексемами природної мови людини; розпізнавання тем текстів, їх класифікація; анотування та реферування текстів; розпізнавання абревіатур; морфологічна декомпозиція (перетворення окремих термінів (наприклад, медичних) у зрозумілу форму); синтез мовлення (озвучення/прочитання тексту); виведення/розпізнавання тексту з відсканованих документів або PDF-файлів; розпізнавання мовлення, продукованого людським голосом; конвертування комп'ютерних даних у природну мову людини та ін.

Отже, в залежності від того, які задачі вирішуються, всі системи, центральним змістом яких є автоматизована обробка текстів можна розділити на класи [1]:

1. Зберігання текстів. Повнотекстові бази даних та інтелектуальний пошук;
2. Системи автоматичного індексування і рубрикації;
3. Системи автоматичного анотування і реферування;

4. Інформаційно-пошукові системи;
5. Системи машинного перекладу;
6. Системи класу «Штучний інтелект» (Текст → База знань);
7. Системи генерації тексту (База знань → Текст).

Очевидно, що потреба в реалізації таких технологій виникла вже давно, з початком розквіту науково-технічного прогресу. Це підтверджується тим фактом, що дослідження з питань обробки природної мови проводяться вже близько 70 років. За цей час дана галузь мала різні назви (що існували паралельно) та визначення, які склалися у наслідок досліджень, що перманентно виконуються та розробок, що постійно реалізуються. Так, обробка природної мови (Natural-language processing, NLP) має такі синоніми як комп'ютерна лінгвістика (Computational Linguistics, CL), технологія людської мови (Human Language Technology, HLT), інженерія природної мови (Natural Language Engineering, NLE) [2]. Найпоширенішим визначенням, на сьогоднішній день, є наступне «Обробка природної мови – один з головних напрямків розвитку інформаційних технологій і комп'ютерних наук на сьогоднішній день щодо вирішення проблем штучного інтелекту та математичної лінгвістики. Воно включає семантико-лінгвістичний, лексикографічний та концептографічний аналіз» [3]. Узагальнено це означає, що штучний інтелект це система, якій притаманна природна людська мова, тобто система розуміє природну мову й також спроможна генерувати смислові природномовні конструкції. Тим самим, можна констатувати, що на сучасному етапі створення систем штучного інтелекту постійно удосконалюється формат взаємодії комп'ютера та людини.

Темі аналізу методів обробки та представлення природномовних текстів присвятили свої роботи такі дослідники, як: Джонс К. [4], Лідді Е. [5], Журафські Д. [2], Барковська О. [6], Превіна С. [7] та ін. Найвідомішу періодизацію наукових пошуків в цій галузі в 2001 році сформувала дослідниця Карен Джонс. Вона виділяє чотири етапи розвитку обробки природних мов:

I етап (кінець 1940-х – кінець 1960-х рр.) акцентування уваги на створенні систем машинного перекладу;

II етап (кінець 1960-х – кінець 1970-х рр.) домінування теорій штучного інтелекту;

III етап (кінець 1970-х – кінець 1980-х рр.) розробка обчислювальних граматики і логічного програмування;

IV етап (1990-ті) застосування статистичного підходу.

Після дослідження К. Джонс пройшло 20 років. Отже, можемо говорити про п'ятий, сучасний етап: домінування експліцитних методів семантичного аналізу текстової інформації (алгоритми онтологічного семантичного аналізу) та методів латентно-семантичного аналізу [8].

Протягом всього зазначеного часу, найбільшими проблемами при аналізі природних мов було визнано: семантичну багатозначність речень природних мов; явище синонімії та омонімії; аналіз емоційного забарвлення фраз; аналіз фраз іронічного характеру; аналіз тексту при неповній інформації, коли мається на увазі щось по-замовчуванню або взагалі невідомо; проблема індивідуальної інтерпретації змісту [1], [9].

Пропонуємо короткий огляд найбільш поширених моделей і методів, що застосовувалися для вирішення окреслених проблем. Зазначимо, що з огляду на значну кількість різних підходів до формалізації природної мови накопичену за кілька десятиліть, подана тут класифікація їх є, до певної міри, умовною [10].

**Обчислювальний підхід.** Першими моделями, які розроблялися для задач автоматичного перекладу були: ієрархія граматики Хомського [11–13], граматики Вудса [14], граматики залежностей [15], ймовірнісні автомати [16], моделі «Зміст-Текст» [17] та ін. Вивчення подібних моделей привело до створення розвиненої теорії формальних мов, в рамках якої були сформульовані чисто математичні задачі, такі як: проблема приналежності слова до мови, заданої формальною граматикою; проблема знаходження перетину двох мов; проблема складності опису мови, тощо.

**Лінгвістичний підхід** – включає чотири технологічні рівня [18]:

графематичний – розмітка та виділення локальних (окремих) часток та елементів текстового масиву / документу / наративу, прикладами яких є розділи, параграфи, абзаци, речення, тощо;

морфологічний – виявлення, виділення, визначення значущих частин окремого слова, значущих слів у реченні, тощо;

синтаксичний – визначення, виявлення логічних схем упорядкування слів у реченні, на основі визначених правил сполучуваності з іншими словами;

семантичний – розуміння смислу текстів.

За іншими джерелами [5], крім названих виділяють також лексичний, контекстний та прагматичний рівні. Лінгвістичний аналіз текстів використовується нині як один з попередніх етапів обробки природномовних документів в більш складних системах.

**Статистичний підхід.** В основі статистичного підходу до обробки природної мови лежить припущення, що зміст тексту може бути визначено за найуживанішими словами. Основним завданням даного підходу є визначення кількості повторень конкретного слова в тексті. Великий внесок в розвиток статистичних методів вніс Клод Шеннон працею «Prediction and entropy of printed English» [19]. Пізніше Кузін Н.Т. [20] описав методи частотної обробки текстової інформації, які згодом були удосконалені в роботах Бродера А. [21] та Ланде Д.В. [22].

**Латентно-семантичний метод** є різновидом статистичного підходу, що реалізується на основі виявлення міри подібності слів, що складають конкретні контексти, при умові що повна множина цих контекстів визначає множину взаємних обмежень щодо виявлення цих подібностей. До статистичного класу методів відносять також умовні випадкові поля. Це роздільні (диференційні) моделі, які формують логістичну регресію для послідовності даних. Дані моделі успішно використовуються для розмічування частин мови, поверхнево-синтаксичного аналізу, розпізнавання іменованих сутностей [23]. Спорідненою з

умовними випадковими полями є прихована марковська модель [24], яка знайшла широке застосування в напрямках розпізнавання мовлення, рукописного введення, морфологічної розмітки, тощо.

**Імовірнісний підхід.** Представлений такими моделями як N-грами, системи засновані на деревах рішень, та імовірнісні узагальнення контекстно-вільних граматики. N-грамна модель реалізується на основі оброблення послідовностей речень, слів, букв, звуків і т.д., виходячи з вибірки із  $n$  елементів, що їх складають. За рахунок накопичення частот появи певних лінгвістичних конструкцій реалізується розрахування ймовірності прояву елементів, що складають текст, який аналізується. Така модель зводиться до скінченної множини ймовірностей, кожна з яких може бути оцінена після обчислення повторюваності відповідних  $n$ -грам. Формальні математичні властивості таких моделей досліджені досить мало. Окремими авторами [25, 26] здійснюються спроби більш детально проаналізувати властивості N-грам, використовуючи для цього апарат теорії автоматів і регулярних мов.

Наступні два підходи сформульовані Елізабет Д. Лідді і подаються за її працею [5].

**Символічний підхід** – реалізується на основі глибинного семантичного аналізу лінгвістичних сутностей, явищ, тощо. Для цього використовуються логістичні схеми представлення знань. Джерелом знання про мову можуть виступати словники, формули та правила, розроблені людьми. Один з прикладів символічного підходу є системи засновані на логіці або на основі правил. У системах, заснованих на логіці, символічна структура зазвичай складається у формі логічних пропозицій. Маніпуляції таких структур визначаються загальнодоступними процедурами збереження правди. Системи, засновані на правилах, зазвичай складаються з набору правил, механізму висновку, робочої області або робочої пам'яті. Знання представлені як факти чи правила. Двигун висновку неодноразово вибирає правило, стан якого задовольняється та виконує правило. Ще один приклад символічних підходів – семантичні мережі. Вперше

запропоновані для моделювання асоціативної пам'яті в психології. Семантичні мережі представляють знання через набір вузлів (об'єктів) та позначених посилань, які представляють відносини між вузлами. Схема взаємозв'язку вузлів відображає смислову організацію тексту. Докладніше про семантичні мережі знайомлять праці [27, 28].

**Коннективістський підхід** – реалізується на основі використання моделей різноманітних лінгвістичних структур, що відображають конкретні приклади лінгвістичних сутностей та мовних явищ. Його характеризує поєднання статистичних методів з різними теоріями, що можуть бути представлені у форматі логічних формул.

Деякі коннективістські моделі називають локалістичними, припускаючи, що кожна одиниця являє собою конкретну концепцію. Наприклад, одна одиниця може представляти поняття «сsaveць» інша одиниця може представляти поняття «кіт». Зв'язки між поняттями закодовані у формі спеціальних вагових коефіцієнтів. Локалістичні моделі досить схожі на семантичні мережі, але зв'язки між одиницями зазвичай позначаються не так, як в семантичних мережах. Локалістичні моделі добре працюють в таких завданнях, як генерування мови та обмежений умовивід. Інші коннективістські моделі називають розподіленими. На відміну від локалістичних моделей, поняття в розподілених моделях представлено як функція одночасної активації кількох одиниць. Ці моделі добре підходять для завдань асоціативного пошуку [29].

Одними з найбільш затребуваних сучасних комп'ютерних систем є знання-орієнтовані системи класу «Штучний інтелект», такі як експертні системи, системи підтримки прийняття рішень та ін. Ними користуються у сфері державного управління, виробництві, бізнесі, науці. Спілкування природною мовою закладено, в деякому вигляді, у вищеназвані системи. Процес аналізу в цих системах повинен завершуватися побудовою семантичної структури, в якій фіксується зміст тексту. Згідно дослідження Леонтєвої Н. Н., можна виділити п'ять різних підходів до побудови такої структури. Ці підходи розширюють уявлення про методи обробки і

представлення природних мов. Пропонується короткий опис існуючих семантичних структур.

**Лінгвістичні структури речень тексту** фіксують результат «буквального» локального (в межах одного речення) розуміння. Лінгвістичні процесори, що спираються на численні ґрунтовні словники тяжіють до повноти інтерпретації кожного конкретного речення вхідного тексту. В основі лінгвістичних моделей лежить синтаксичне (синтаксично-семантичне) представлення речення. Класичні лінгвістичні структури – це структури, в основі яких лежать лінгвістичні моделі. Серед переваг лінгвістичних структур: детальність аналізу, що відображається у формі синтаксичного або семантичного дерева, а також можливість спів ставляти будь-якому реченню вхідного тексту його формальний структурний образ. Серед недоліків: жорсткі рамки лінгвістичних одиниць, в яких можна моделювати лише розуміння в межах речення без узагальнень та слабка кореляція з одиницями представлення знань.

**Семантична мережа цілого тексту.** Це структури, в яких зафіксовано відомості про тему-рему, емпатичні акценти, або встановлено хоча б референтні зв'язки між структурами сусідніх речень. Після перекладу синтактико-семантичного представлення всіх речень на мову більш елементарних одиниць, отримують мережу, глобальну, «розміту» структуру. В таких структурах вузли більш змістовні та експліцитні ніж зв'язки, і тому, інколи, робота з такими моделями є не досить ефективною, тому що в глобальному семантичному просторі цілого тексту діють інші закони, ніж в межах речення.

**Інформаційні структури цілого тексту** фіксують результат глобального розуміння тексту і потоку текстів в одиницях термінології обраної предметної галузі. Терміни зосереджені в джерелах, що задаються окремо від тексту: класифікаторах, тезаурусах, рубрика торах і т.д. Системи автоматичного індексування та інформаційно-пошукові системи, що працюють з такими структурами, мають справу з реальними текстовими масивами. Перевагами даного виду структур є: можливість обробки текстів без будь-яких структурних обмежень;

центральними елементами є терміни, а вони, як правило, не такі дрібні як в лінгвістичних моделях, і не такі великі, як в системах штучного інтелекту, тобто вони «середні» і мають еквіваленти для перекладу їх іншими мовами.

**Структури баз даних та знань** – це формальні, жорсткі, фіксовані структури, які відображають частину дійсності (наприклад, таблиці з описом товарів виробництва), які можна застосовувати при аналізі природномовного тексту як додаткові знання. Над цими структурами можливе виконання математичних операцій, що, безперечно є перевагою. Серед даних моделей є і напівжорсткі структури «динамічного» типу, це – сценарії, схеми, або їх частини – фрейми. Такі структури отримали широке застосування в системах класу ШІ, вони відображають сюжет цілого тексту, який, як відомо, байдужий до членування на речення. При використанні їх в якості інструмента аналізу, схема сюжету, що жорстко задана, доповнюється лінгвістичними специфікаціями, що дозволяють виокремлювати з тексту і вносити до бази знань інформацію, якої бракує. Перевагою даних структур вважається можливість порівняння змісту тексту що аналізується, з різними джерелами інформації, зокрема з іншими текстами тієї ж тематики. Семантичні структури, які прагнуть побудувати подібні моделі, називаються концептуальними структурами.

**Структури систем машинного перекладу** реалізують найповніший ланцюг роботи з текстом: від вхідного тексту до вихідного, який належить до іншої мови. Вони фіксують лексичні та синтаксичні відповідності (й розходження) між одиницями і структурами двох мов. Ще більше розходження між вхідними та вихідними одиницями можна спостерігати в структурах інформаційно-перекладних систем, якщо вони здійснюють переклад зі стисненням, спираючись на бази знань: одиниці вхідного та вихідного текстів належать різним природним мовам, відображають різні обсяги змісту, можуть різнитися фокусами уваги. Такого роду системи автоматичного перекладу, засновані на знаннях (knowledge based machine translation systems), які поєднують в собі повноту лінгвістичного



аналізу зі структурами знань є найбільш перспективними системами автоматичного розуміння тексту.

Відносно новим класом комп'ютерних систем вважаються *знання-орієнтовані інформаційні системи з онтолого-керованою архітектурою* [30, 31]. Вони реалізують інтегровану інформаційну технологію, що включає комп'ютерну обробку природномовних об'єктів, заданих лінгвістичним корпусом текстів, які описують деяку проблемну галузь, та вилучення предметно-орієнтованих знань з метою їх формально-логічного представлення та автоматизованої обробки. Вивченню і вдосконаленню знання-орієнтованих систем, в тому числі з онтолого-керованою архітектурою, присвятили свої праці такі вчені, як Глушков В. М. [32, 33], Гладун В. П. [34, 35], Палагін О. В. [9, 30], Широков В. А. [36], [37], Хорошевський В. Ф. [38], Поспелов Д. А. [39], Овдій О. М. [40], Гаврилова Т. А. [41, 42], Загорулько Ю. А. [43], Малишевський А. В. [44], Андон П. І. [45], Валькман Ю. Р. [46], Стрижак О. Є. [47–49], Величко В. Ю. [50], Gruber T. [51], Guarino N. [52, 53], Noy N. [54], Gomez-Perez A., Fernandez Lopez M., Corcho O., [55] Gruninger M. [56], Happel H. [57], Mc Guinness D. [58].

Даний клас інформаційних технологій виник внаслідок інтеграції двох галузей ШІ, які раніше розвивалися паралельно: knowledge-engineering та комп'ютерна лінгвістика. Цьому сприяла низка факторів, таких як:

- удосконалення методів комп'ютерної обробки природномовної інформації та предметних знань;
- системна інтеграція міждисциплінарних знань як таких;
- закономірний еволюційний розвиток наукових теорій, закладених ще з давніх-давен, відомих як логіка та онтологія.

Не дивлячись на широку популярність онтологічного інжинірингу серед фахівців з комп'ютерних наук, і, як наслідок, відчутну кількість фундаментальних і прикладних досліджень в цій галузі, вона залишається актуальною і перспективною. Пояснення цього факту подає Палагін О. В.: «Наявний в даний час розрив між добре опрацьованими методами і засобами обробки природномовної

інформації для вирішення прикладних завдань в вузькоспеціалізованих предметних галузях, з одного боку, і недостатністю таких стосовно до вирішення комплексних завдань, пов'язаних з аналізом і розумінням природномовної інформації, її формально-логічним поданням, витяганням предметних знань з їх подальшим використанням в загальнозначущих предметних галузях, в тому числі і для проведення складних міждисциплінарних наукових досліджень, стримує розвиток і застосування систем обробки знань, що реалізують (в тому числі) в повному обсязі технології Data & Text Mining. Особливо слід відзначити дефіцит якісних методів і засобів знання-орієнтованої обробки природномовної інформації для україномовного сегменту» [30].

### **1.1.2. Моделі представлення природномовної інформації**

Аналіз природномовних текстів неминує пов'язаний з необхідністю представляти в деякий спосіб результати такого аналізу. Оскільки аналіз завжди проводиться з якоюсь метою, то інформація, вилучена з тексту буде мати статус знань [59, 60]. Проблема представлення знань актуальна для таких наукових галузей, як когнітивна психологія, штучний інтелект та розробка експертних систем.

Людина одночасно використовує найрізноманітніші методи представлення знань: мовний опис, графічну інформацію, математичні формули, тощо. Слід відмітити, що функціональна специфіка діяльності довільного експерта визначає види та типи інформаційних структур, наративів, документів та їх контекстів, що він використовує. Так, філологи працюють над текстовою інформацією; математики, фізики, в більшості, опрацьовують формальні математичні нотації з різних теорій; культурологи, мистецтвознавці опрацьовують, у більшості, гібридний формат графічної та текстової інформації, де центр сенсу зміщено до графічних образів. [61].

Тому, й про це свідчить сучасний розвиток інформаційних технологій, універсально визначити штучно створену універсальну систему представлення знань практично неможливо. Необхідна досить велика зв'язність базових елементів

кожного виду і типу інформації та даних, що не може бути реалізована сучасними апаратними і програмними засобами. А також необхідно дану зв'язність функціонально інтерпретувати. Тому, найбільш ефективно дану проблему вирішувати з використанням різних методів представлення знань, кожен з яких є найбільш раціональним для тієї чи іншої предметної галузі [41].

До їх переліку класично слід віднести наступні формати представлення семантичного рівня знань: предикати першого порядку, продукційні правила, семантичні мережі, фрейми, нейронні мережі, онтології. Використання кожного з них при реалізації прикладних систем обумовлене функціональною специфікацією конкретної предметної галузі. Також ефективність таких рішень суттєво залежить від рівня технічних і програмних засобів, що використовуються в задачах даної предметної галузі.

Що стосується моделей представлення знань, то їх велику кількість класифікують за ступенем декларативності мовних засобів, що використовуються для опису знань, на декларативні та процедуральні.

Декларативна модель представлення знань реалізується на основі статичного представлення структури та на її основі логістичних взаємозв'язків представлення описів елементів, понять та їх властивостей. Це дозволяє використовувати ці моделі знань у різних задачах. Однак, ця їх структурна інваріантність не досить продуктивна у процесах глибинного семантичного аналізу.

На відміну від декларативної, процедуральні моделі знання реалізуються у форматі спеціальних процедур – певних програм, які реалізують генерацію логістичних взаємозв'язків на основі оброблення властивостей понять, що змістовно складають систему знань. За рахунок цього забезпечується досягнення мети, що визначається конкретною задачею. Тобто процедуральна модель несе в собі набір цілей, які треба досягнути й правила у форматі процедур, які реалізують генерацію логістичних схем досягнення визначених цілей. Тому процедурна модель виключає необхідність описування усіх станів рішення задачі. Логістичні схеми, що генеруються на визначених умовах, забезпечують виводимість рішення.

Функціональна різниця між декларативним і процедурним представленням систем знань полягає у наступному: декларативна забезпечує відображення того «що ми знаємо», а процедурна «як це можна використати». Це в свій час породило напрямок створення гібридних систем, які забезпечили визначення над структурними представленнями знань, певних функцій, які реалізовували генерацію відповідних логістичних схем рішення задачі на основі оброблення властивостей понять, що представлені у декларативних форматах. Сюди відносяться фрейми, онтології, нейромережі, тощо.

Також моделі представлення знань поділяють на логічні та евристичні.

Логічні моделі представлення знань засновані на понятті формальної системи. Їхніми прикладами можуть бути числення предикатів і будь-яка конкретна система продукцій. Вони можуть бути дедуктивними чи індуктивними. Дедуктивне виведення рішення забезпечується певними наборами правил, які визначають первинні умови задачі та процес оброблення елементів знань. Індуктивні моделі є узагальненням дедуктивних. Вони реалізують процес виведення рішення на основі оброблення кінцевого числа навчаючих прикладів.

Евристичні моделі представлення знань, на відміну від логічних, мають різноманітний набір засобів, що передають специфічні особливості моделі. Ці особливості у форматі певних ознак, використовуються для генерації за певними правилами схеми виводимості рішення задачі. Вони носять імовірнісний характер тому забезпечують нечітке виведення рішення. До евристичних моделей варто віднести мережні, фреймові і продукційні моделі.

Жодна з моделей подання знань не є лідером за всіма критеріями, які висуваються до них [28].

**Продукційні моделі** [34, 38] (або моделі засновані на правилах) описують процедурні знання і дозволяють представити їх у вигляді пропозицій типу «якщо (умова) ..., то (дія)». Механізм, реалізований як засіб виводу в продукційних системах, називається машиною логічного висновку і виконує функції пошуку в базі правил, послідовно виконуючи операції над знаннями та отримуючи висновки.

Існує два способи отримання таких висновків – прямі висновки і зворотні. Прямі висновки виводять від посилок до слідств.

Зворотні висновки реалізуються на основі формування спочатку мети задачі, й передумов її розв'язання. Умова визначається формалізованим представленням зразка, який забезпечує пошук у базі знань (фактів). Звичайно такий пошук реалізується методом резолюцій, який забезпечує реалізацію проміжних та кінцевих логістичних схем знаходження та виявлення фактів. Кінцеві виведення є термінальними, визначення яких фіксує розв'язання задачі. При формуванні логістичних схем виникають проміжні факти, які виявляються у станах виявлення допоміжних фактів, що сприяють формуванню ознак фактів, які треба виявити. Тобто база знань у форматі продукційної моделі являє собою зв'язну сукупність фактів і правил.

До переваг продукційних моделей можна віднести: наочність і зрозумілість знань (принаймні, на рівні одного правила); можливість реалізації немонотонного логічного висновку та обробки суперечливих фактів; можливість введення різних модифікацій в інтерпретацію правил відповідно до особливостей розв'язуваних системою завдань; можливість легкого нарощування бази знань шляхом додавання нових правил. Серед недоліків: неосязність великої бази знань і її структури; можливість легкого внесення серйозних помилок в базу знань, що призводять до неправильного функціонування системи (якщо в системі немає розвинених засобів перевірки цілісності бази знань); орієнтація на послідовну обробку правил.

**Семантичні мережі** [27]. Це системи представлення знань у форматі графових структур за допомогою вузлів, що з'єднуються дугами. Це можуть бути орієнтовані граfi та орієнтовані дерева. У таких структурах вузли є поняттями певних предметних галузей, а дуги фіксують певні відношення між ними.

Семантичні мережі досить ефективно реалізують асоціативні зв'язки між поняттями та їх класами. Це досить суттєво технологічно сприяє індуктивному виведенню рішень. У такому випадку семантично визначаються смисли об'єктів та їх відношень. В якості понять виступають як абстрактні так і конкретні об'єкти, а

основу відношень складають зв'язки типу: «це», «має частиною», «належить» і т.д. Характерною рисою семантичних мереж є наявність трьох типів відносин: 1) клас – елемент класу (квітка – троянда); 2) властивість – значення (колір – жовтий); 3) приклад елемента класу (троянда чайна).

Семантичні мережі мають певний рівень універсальності, який забезпечується множиною визначених відношень. Вони мають свої прояви при формуванні структурних описів різних ситуацій, станів, тощо у довільній предметній галузі.

Однак їх практична незорість при відображенні великих обсягів інформації суттєво заважає їх широкому застосуванню.

**Фрейми** [62] Фреймова модель є систематизованою структурною моделлю пам'яті людини і її свідомості. Досить конструктивно цей формат моделі системи знань наведено у Вікіпедії: «Фрейм (англ. *frame* – рамка, каркас) це структура даних для представлення деякого концептуального об'єкта. Це мінімальна структура інформації, необхідна для представлення класу об'єктів, явищ або процесів. Інформація, що відноситься до фрейму, міститься в слотах, що його складають. Слот – це деякі незаповнені підструктури фрейму, заповнення яких приводить до того, що цей фрейм ставиться у відповідність деякій ситуації, події або об'єкту. Слот може бути термінальним (містити тільки ім'я слота і значення імені) або являти собою фрейм нижнього рівня. У вигляді фрейму може описуватися деякий об'єкт, ситуація, абстрактної поняття, формула, закон, правило, візуальна сцена і т. ін. Особливістю і основним недоліком фреймового представлення знань є відсутність механізму висновку, тобто відповідальність за правильну обробку знань повністю покладається на розробника системи, який повинен представляти предметну галузь і не тільки детально описати кожен фрейм і слот, але і правильно задати процедури їх обробки. До основних процедур управління висновком у фреймовій моделі звичайно відносять: механізм успадкування, обмін повідомленнями між фреймами, приєднані процедури» [63].

**Формальні логічні моделі** [39, 64, 65] передбачають, що вся інформація, необхідна для вирішення прикладних завдань, розглядається як сукупність фактів

і тверджень, які представляються як формули в деякій логіці. Знання відображаються сукупністю таких формул, а отримання нових знань зводиться до реалізації процедур логічного висновку. Логічні методи забезпечують розвинутий апарат виведення нових фактів на засадах композиції вже визначених у базі знань фактів.

Основним правилом оброблення таких форматів знань є логічне виведення. Для їх створення існує певна множина логічних мов, які ґрунтуються на засадах теорії предикатів першого порядку та численні висловлювань.

Перевагами логічного підходу є: наявність логічних формалізмів у форматі предикативного представлення фактів у базі знань; високий рівень модульності знань і одночасно можливість одержання єдиної системи представлення системи знань. Однак, слабе використання конструкцій природної мови у процесі визначення фактів у базі знань (у зв'язку з тим, що механізми формування лінгвістичних конструкцій, через відсутність формальної семантики природної мови не розвинутий), а також відсутність чітких принципів організації фактів у базі знань та надмірний рівень формалізації подання знань суттєво знижує ефективність цього підходу, й робить його неможливим для оброблення великих обсягів інформації.

## **1.2. Трансдисциплінарність як критерій формування сучасних систем обробки та представлення природномовних текстів**

Поняття трансдисциплінарності увійшло в науково-інформаційний обіг на початку 70-х років ХХ ст. і позиціонує себе як нетривіальний спосіб наукового мислення, що полягає в ідеї розгляду будь-якого феномену поза рамками конкретних наукових дисциплін і покликаний запропонувати ефективні методи рішення складних викликів сучасності. Даний підхід, запропонований Ж. Піаже, осмислювався багатьма дослідниками (від філософів до математиків), в результаті чого виникло декілька альтернативних трактувань поняття трансдисциплінарності. Серед них виділяють такі:

– в першому значенні трансдисциплінарність розуміється як «декларація», що проголошує рівні права відомих і маловідомих вчених, більших і менших наукових дисциплін, культур і релігій в дослідженні навколишнього світу;

– в другому значенні трансдисциплінарність трактується як високий рівень образності, різнобічності, універсальності знань конкретної людини;

– в третьому значенні трансдисциплінарність трактується як «правило дослідження навколишнього світу». Передбачається, що трансдисциплінарність буде реалізована, якщо проблема одразу досліджується відповідно до вимог всебічності. Наприклад, на фізичному, соціальному і психічному рівнях, як локальне, в якому глобальне;

– у четвертому значенні трансдисциплінарність використовується як «принцип організації наукового знання», що передбачає взаємодію багатьох напрямків наукового дослідження при вирішенні комплексних проблем розвитку людини, природи і суспільства. Слід зазначити, що трансдисциплінарність в цьому сенсі передбачає вихід вчених за рамки своїх дисциплін.

Докладніше різні трактовки і підходи до розуміння трансдисциплінарності, а також виділення її видів і можливих перспектив розвитку, висвітлено в роботах Клейна Дж. [66], Ніколеску Б. [67], Джаджа Е. [68], Князевої О. [69], Кіященко Л. [70], Лоуренса М. [71], Ренна О. [72] та ін.

Узагальнюючи ці підходи, можна сказати, що трансдисциплінарність означає у вузькому сенсі інтеграцію різних форм і методів дослідження наукових проблем, а у широкому – єдність знання за межами конкретних дисциплін [73]. Саме це, широке розуміння трансдисциплінарності і цікавить нас в сенсі можливого шляху вирішення деяких проблем у сфері інформаційних технологій.

Однією з головних переваг сучасної концепції трансдисциплінарності є те, що вона дає можливість досліднику об'єктивно розглядати всі аспекти явища, що вивчається. Постулювання «єдності світу» дозволяє розглядати досліджувані об'єкти або процеси як природний фрагмент глобального середовища, тобто розглядати їх в образі «єдиного природного функціонального ансамблю». Таким



чином, методологія трансдисциплінарності – це спосіб розуміння, пізнання і опису об'єкта у складі Єдиного впорядкованого середовища [74].

Трансдисциплінарність дозволяє розглядати усі процеси в інформаційному просторі на основі певної множинної упорядкованості станів взаємодії систем, які її складають. А це означає, що до вирішення таких проблем, як: неупорядкованість великих масивів інформаційних ресурсів в глобальному середовищі; забезпечення процесу коректного одночасного використання мережевих інформаційних ресурсів, що обробляються різними інформаційними системами і які не мають спільних інтерфейсів; побудова консолідованого середовища взаємодії мережевих ресурсів (особливо актуальних при роботі експертних груп під час розв'язання складних, багатоаспектних проблем); слід підходити з позиції трансдисциплінарності. Конструктивно, така консолідована взаємодія інформаційних систем з використанням мережевих інформаційних ресурсів можлива, якщо ці системи умовно розмістити у єдиному і певним чином упорядкованому інформаційному просторі [74].

Шляхом до створення такого простору є конструювання концептуально-понятійного каркасу наукових теорій, в якості якого може служити сукупність формальних комп'ютерних онтологій конкретних предметних галузей [75]. Тут вважаємо за доцільне навести схему такої конструкції, розроблену Стрижаком О. Є., яка є найбільш наочною (рис. 1.1).

Центральне місце в побудові такої структури посідає створення (як експертне, так і автоматизоване) онтологій. Онтологія [54, 56, 76–78], це представлення деякої предметної галузі (ПГ) за допомогою концептуальної схеми, що зазвичай складається з ієрархічної структури даних, їх зв'язків і правил, прийнятих в цій галузі.

Візуальні способи представлення інформації, а саме онтології, таксономії, класифікатори, семантичні мережі, тезауруси, словники можуть максимально використовувати весь лінгвістичний апарат тієї чи іншої мови.

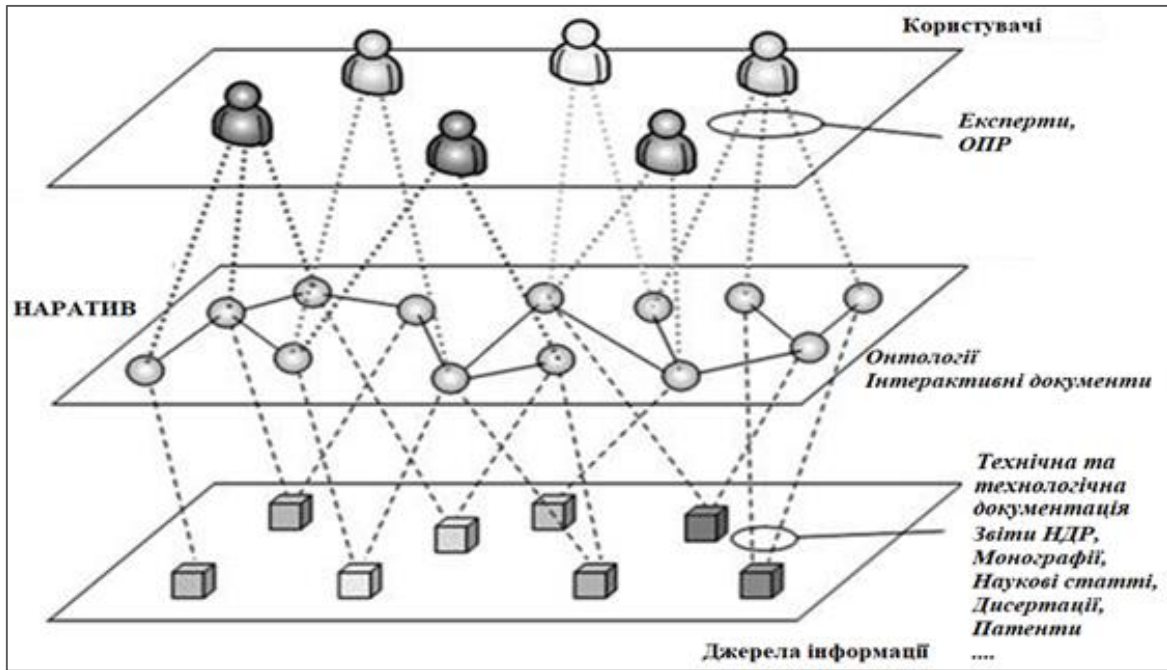


Рисунок 1.1 - Узагальнена структура взаємодії, на засадах спільного використання єдиного наративу інформаційних ресурсів

Актуальність онтологічного підходу до представлення ПГ визначається двома факторами. По-перше, формальні комп'ютерні онтології є одночасно і результатом розвитку й інструментом knowledge engineering. По-друге, вони виконують функції концептуалізації та специфікації наукових теорій, будучи основою побудови сучасних інформаційних технологій [75]. Онтологічний підхід надає користувачам цілісний, системний погляд на визначену ПГ, інформаційні джерела про ПГ представлені однотипово, що спрощує їх сприйняття, а також побудова онтологій дозволяє відновити логічні зв'язки ПГ, яких не вистачає [79]. В онтолого-класифікаційній схемі засобів і методів ШІ онтологічний підхід трактується як різновид системного підходу, заснованого на формуванні знань.

Онтологічні аспекти трансдисциплінарної інтеграції інформаційних ресурсів викладено в роботі [80], і полягають вони в наступному: «...трансдисциплінарні онтології забезпечують коректне агрегування різних тематичних процесів шляхом формування структурованої сукупності інформаційних об'єктів-концептів ПГ, що визначаються як єдиний тип даних. Технологія їх використання в мережевому середовищі, в якому активуються процеси взаємодії складних інформаційних

систем, дозволяє встановити над інформаційними ресурсами, що активно використовуються відношення часткового порядку. Розгляд інформаційних ресурсів як тематичних систем знань, дозволяє визначити їх семантичні характеристики на основі виділення інформаційних одиниць у вигляді концептів. Трансдисциплінарний розгляд контекстів цих одиниць-концептів забезпечує їх консолідоване використання в процесі вирішення складних прикладних задач. Одним з конструктивних способів інтеграції інформаційних ресурсів, як пасивних систем знань є активізація їх концептів на основі формування з них тематичних онтологій і об'єднання цих онтологій на основі побудови над ними онтології задачі вибору».

Інакше кажучи, трансдисциплінарність ґрунтується на встановленні формального взаємозв'язку розумінь синергії окремих тематичних предметних галузей. Вона забезпечує формування логічних мета-рамок, за допомогою яких знання можуть бути інтегровані на більш високому рівні абстракції, ніж це відбувається в міждисциплінарності.

### **1.3. Порівняння існуючих систем обробки та представлення природномовних текстів**

При порівнянні систем, які використовуються для обробки масивів інформації, що містять неструктуровані природномовні тексти, за мету ставиться виявлення серед них систем, що не тільки проводять глибокий лінгвістичний аналіз, але й здатні представляти результати такого аналізу у вигляді, придатному застосовувати їх в подальшому процесі трансдисциплінарної консолідації інформаційних ресурсів. Адже, як зауважив Когаловський М. в своїй праці «Перспективні технології інформаційних систем»: «... поряд з „чистими“ системами... створюються численні крупні системи, що сумісно використовують різноманітні базові технології» [81], що, очевидно, є ознакою їх інноваційності та розвиненості, що важливо для кінцевого споживача. В свою чергу, якщо така система має технологічні можливості для вбудовування її у ще більш складну систему – це

підвищує її цінність для інженерів інформаційних технологій. Також враховувався фактор придатності систем для обробки україномовних ресурсів.

Методика порівняння існуючих систем обробки та представлення природномовних текстів базується на процедурі ранжування альтернатив [82] методом багатокритеріальної оптимізації. Для реалізації поставленої задачі необхідно:

- 1) Відібрати системи для аналізу;
- 2) Сформулювати критерії якості систем;
- 3) Провести оцінку обраних систем за розробленими критеріями;
- 4) Визначити важливість кожного з критеріїв для реалізації поставленої задачі;
- 5) Виконати ранжування систем і вибрати систему з найвищою оцінкою.

Для аналізу обрано 26 систем, серед яких представлено системи для обробки природномовних текстів українською, російською, англійською та іншими мовами; охоплено системи різних рівнів лінгвістичного аналізу; доступні безкоштовно і комерційні, а саме: Поліедр-Рекурсивний редуктор; General Architecture for Text Engineering; Affx Grammars Over a Finite Lattice (AGFL); RapidMiner; LexicoSyntactic Pattern Language (LSPL); Тоμίта-парсер; Конспект (TextTermin); ИСИДА-Т; Link Grammar Parser; Link Grammar Parser for Russian; RCO (Russian Context Optimizer); libmorphukr; Mystem; Cibola/Oleada; pymorphy2; Russian Morphological Dictionary; Stemka; MonoConc; TextAnalyst 2.0; AskNet; Vivisimo; Quintura Searchcrystal; МедіаЛінгва; ABBYY Retrieval & Morphology (ARM) Engine; Apache Lucene/Solr; Convera RetrievalWare.

Серед обраних зразків представлені системи різних класів, а саме:

– *Лінгвістичні фреймворки* (Поліедр-Рекурсивний редуктор; General Architecture for Text Engineering; Affx Grammars Over a Finite Lattice (AGFL); RapidMiner; LexicoSyntactic Pattern Language (LSPL); Тоμίта-парсер). Є найбільш розвиненими системами за кількістю наявних функцій, а також допускають можливість модифікації програмного засобу.

– **Синтаксичні аналізатори** (Конспект (TextTermin); ИСИДА-Т; Link Grammar Parser; Link Grammar Parser for Russian; RCO (Russian Context Optimizer)). Це програмний компонент, який приймає вхідні дані у вигляді ПМТ і створює структуру даних (часто дерево розбору), абстрактне дерево синтаксису або іншу ієрархічну структуру, тим самим забезпечуючи структурне представлення вводу, а також перевіряє правильність синтаксису.

– **Морфологічні аналізатори** (libmorphukr; Mystem; Sibola/Oleada; rymorphy2; Russian Morphological Dictionary; Stemka). Містять велику кількість лінгвістичної інформації, збір і упорядкування якої є трудомістким процесом. Тому вони ціняться як готові бази мовних даних.

– **Пошукові машини** (MonoConc; TextAnalyst 2.0; AskNet; Vivisimo; Quintura Searchcrystal; МедіаЛінгва; ABBYY Retrieval & Morphology (ARM) Engine; Apache Lucene/Solr; Convera RetrievalWare). Деякі пошукові машини мають базові можливості по морфологічному аналізу слів, а також розвинені механізми зчитування ПМТ. Це дозволяє проводити вибірку даних.

Критеріями оцінки систем пропонується вважати наявність чи відсутність в них певних функцій, релевантних рішення поставленої задачі. Оцінювати критерії пропонується по шкалі від нуля до десяти, де нуль – найгірше значення, десять – найкраще. Всі критерії поділяються на шість категорій, зокрема такі:

Підтримка мов. Є критично важливою функцією для систем, призначених для роботи на території України. Результати оцінки наведено на рис 1.2.

№	Системи	Підтримка мов			№	Системи	Підтримка мов		
		укр.	рос.	англ.			укр.	рос.	англ.
1	Рекурсивний редуктор General Architecture for Text	10	10	10	14	Cibola/Oleada	0	0	10
2	Engineering	0	0	10	15	rumorphy2	0	10	0
3	Affx Grammars Over a Finite Lattice (AGFL)	0	0	10	16	Russian Morphological Dictionary	0	10	0
4	RapidMiner	0	0	10	17	Stemka	0	10	0
5	LexicoSyntactic Pattern Language (LSPL)	0	0	10	18	MonoConc	0	0	10
6	Томіта-парсер	10	10	0	19	TextAnalyst 2.0	0	10	0
7	Конспект (TextTermin)	10	10	10	20	AskNet	10	10	10
8	ИСИДА-Т	0	10	0	21	Vivisimo	0	0	10
9	Link Grammar Parser	0	0	10	22	Quintura Searchcrystal	0	0	10
10	Link Grammar Parser for Russian	0	10	0	23	МедіаЛінгва ABBYY Retrieval & Morphology (ARM)	0	10	0
11	RCO (Russian Context Optimizer)	0	10	0	24	Engine	0	10	0
12	libmorphukr	10		0	25	Apache Lucene/Solr	10	10	10
13	Mystem	10	10	0	26	Convera RetrievalWare	10	10	10

Рисунок 1.2 – Результати оцінки за категорією "Підтримка мов"

*Вхідні формати файлів.* Аналізувалася підтримка таких поширених форматів як doc/docx, xls, xlsx, txt, pdf. Результати представлені на рис. 1.3.

№	Системи	Вхідні формати файлів				№	Системи	Вхідні формати файлів			
		doc, docx	xls, xlsx	txt	pdf			doc, docx	xls, xlsx	txt	pdf
1	Рекурсивний редуктор General Architecture for Text	10	10	10	10	14	Cibola/Oleada	0	0	10	0
2	Engineering	0	0	10	0	15	rumorphy2	0	0	10	0
3	Affx Grammars Over a Finite Lattice (AGFL)	0	0	10	0	16	Russian Morphological Dictionary	0	0	10	0
4	RapidMiner	0	0	10	0	17	Stemka	0	0	10	0
5	LexicoSyntactic Pattern Language (LSPL)	0	0	10	0	18	MonoConc	0	0	10	0
6	Томіта-парсер	0	0	10	0	19	TextAnalyst 2.0	0	0	10	0
7	Конспект (TextTermin)	0	0	10	0	20	AskNet	10	0	10	10
8	ИСИДА-Т	0	0	10	0	21	Vivisimo	10	0	10	10
9	Link Grammar Parser	0	0	10	0	22	Quintura Searchcrystal	10	0	10	10
10	Link Grammar Parser for Russian	0	0	10	0	23	МедіаЛінгва ABBYY Retrieval & Morphology	10	0	10	10
11	RCO (Russian Context Optimizer)	0	0	10	0	24	(ARM) Engine	10	0	10	10
12	libmorphukr	0	0	10	0	25	Apache Lucene/Solr	10	0	10	10
13	Mystem	0	0	10	0	26	Convera RetrievalWare	10	0	10	10

Рисунок 1.3 – Результати оцінки за категорією "Вхідні формати файлів"

*Вихідні формати файлів.* Формування вихідних файлів в одному з заданих форматів (HTML, XML PDF, JSON, TXT) має важливе значення як для відображення користувачу, так і для подальшої роботи експертів. Результати приведені на рис. 1.4.

№ Системи	Вихідні формати файлів					№ Системи	Вихідні формати файлів				
	XML	HTML	PDF	JSON	TXT		XML	HTML	PDF	JSON	TXT
1 Рекурсивний редуктор General Architecture for Text Engineering	10	10	10	10	10	14 Cibola/Oleada	0	0	0	0	10
2 Affx Grammars Over a Finite Lattice (AGFL)	0	0	0	0	10	15 pymorphy2	0	0	0	0	10
3 RapidMiner	0	0	0	0	10	16 Russian Morphological Dictionary	0	0	0	0	10
4 LexicoSyntactic Pattern Language (LSPL)	0	0	0	0	10	17 Stemka	0	0	0	0	10
5 Томіта-парсер	0	0	0	0	10	18 MonoConc	0	0	0	0	10
6 Конспект (TextTermin)	0	0	0	0	10	19 TextAnalyst 2.0	0	0	0	0	10
7 ИСИДА-Т	0	0	0	0	10	20 AskNet	0	10	0	0	0
8 Link Grammar Parser	0	0	0	0	10	21 Vivisimo	0	10	0	0	0
9 Link Grammar Parser for Russian	0	0	0	0	10	22 Quintura Searchcrystal	0	10	0	0	0
11 RCO (Russian Context Optimizer)	0	0	0	0	10	23 МедіаЛінгва ABBYY Retrieval & Morphology (ARM)	0	10	0	0	0
12 libmorphukr	0	0	0	0	10	24 Engine	0	10	0	0	0
13 Mystem	0	0	0	0	10	25 Apache Lucene/Solr	0	10	0	0	0
						26 Convera RetrievalWare	0	10	0	0	0

Рисунок 1.4 – Результати оцінки за категорією "Вихідні формати файлів"

Доступні види аналізу (семантичний, синтаксичний, морфологічний, аналіз розмітки, аналіз стилю, аналіз метаданих) визначають спектр інформації що може бути виділена з інформаційних ресурсів. Результат оцінки наведено на рис. 1.5.

№ Системи	Доступні види аналізу						№ Системи	Доступні види аналізу					
	семантичний	синтаксичний	морфологічний	аналіз розмітки	аналіз стилю	аналіз метаданих		семантичний	синтаксичний	морфологічний	аналіз розмітки	аналіз стилю	аналіз метаданих
1 Рекурсивний редуктор	10	10	10	10	10	10	14 Cibola/Oleada	0	0	10	0	0	0
2 General Architecture for Text Engineering Affx Grammars Over a Finite Lattice (AGFL)	10	10	10	10	0	0	15 pymorphy2	0	0	10	0	0	0
3 RapidMiner	10	10	10	10	0	0	16 Russian Morphological Dictionary	0	0	10	0	0	0
4 LexicoSyntactic Pattern Language (LSPL)	10	10	10	10	0	0	17 Stemka	0	0	10	0	0	0
5 Томіта-парсер	10	10	10	0	0	0	18 MonoConc	0	0	10	0	10	10
6 Конспект (TextTermin)	2	10	10	0	0	0	19 TextAnalyst 2.0	0	0	10	0	10	10
7 ИСИДА-Т	0	10	10	0	0	0	20 AskNet	0	0	10	0	10	10
8 Link Grammar Parser	0	10	10	0	0	0	21 Vivisimo	0	0	10	0	10	10
9 Link Grammar Parser for Russian	0	10	10	0	0	0	22 Quintura Searchcrystal	0	0	10	0	10	10
10 RCO (Russian Context Optimizer)	0	0	10	0	0	0	23 МедіаЛінгва ABBYY Retrieval & Morphology (ARM)	0	0	10	0	10	10
11 libmorphukr	0	0	10	0	0	0	24 Engine	0	0	10	0	10	10
12 Mystem	0	0	10	0	0	0	25 Apache Lucene/Solr	0	0	10	0	10	10
							26 Convera RetrievalWare	0	0	10	0	10	10

Рисунок 1.5 – Результати оцінки за категорією "Доступні види аналізу"

Також системи оцінювалися за більш спеціалізованими критеріями, зв'язаними з особливостями представлення результату. Це такі категорії як:

*Можливість виділення з тексту* таких об'єктів, як терміни, зв'язки, атрибути, контексти та зображення. Це визначає, наскільки повно буде представлена наявна у виданні інформація в результаті структуризації. Таблиця оцінок на рис. 1.6.

№	Системи	Можливість виділення з тексту					№	Системи	Можливість виділення з тексту				
		термінів	зв'язків	атрибутів	контекстів	зображень			термінів	зв'язків	атрибутів	контекстів	зображень
1	Рекурсивний редуктор	10	10	10	10	10	14	Cibola/Oleada	10	0	0	0	0
2	General Architecture for Text Engineering	10	10	10	10	0	15	rumorphy2	10	0	0	0	0
3	Affx Grammars Over a Finite Lattice (AGFL)	10	10	10	10	0	16	Russian Morphological Dictionary	10	0	0	0	0
4	RapidMiner	10	10	10	10	0	17	Stemka	10	0	0	0	0
5	LexicoSyntactic Pattern Language (LSPL)	10	10	10	10	0	18	MonoConc	10	0	0	0	0
6	Томіта-парсер	10	10	10	10	0	19	TextAnalyst 2.0	10	0	0	0	0
7	Конспект (TextTermin)	10	10	10	10	0	20	AskNet	10	0	0	0	0
8	ИСИДА-Т	10	10	10	10	0	21	Vivisimo	10	0	0	0	0
9	Link Grammar Parser	10	10	10	10	0	22	Quintura Searchcrystal	10	0	0	0	0
10	Link Grammar Parser for Russian	10	10	10	10	0	23	МедіаЛінгва	10	0	0	0	0
11	RCO (Russian Context Optimizer)	10	10	10	10	0	24	ABBYY Retrieval & Morphology (ARM) Engine	10	0	0	0	0
12	libmorphukr	10	0	0	0	0	25	Apache Lucene/Solr	10	0	0	0	0
13	Mystem	10	0	0	0	0	26	Convera RetrievalWare	10	0	0	0	0

Рисунок 1.6 – Результати оцінки за категорією "Можливість виділення з тексту об'єктів"

Представлення обробленого тексту. Спосіб представлення обробленого тексту (у вигляді тексту, веб-документу, онтології, інтерактивного документу) визначає, чи потрібно буде застосовувати допоміжне програмне забезпечення для відображення результатів структуризації видань. Результати оцінки на рис. 1.7.

№	Системи	Представлення обробленого тексту					№	Системи	Представлення обробленого тексту				
		текст	веб-документ	онтологія	інтерактивний документ	документ			текст	веб-документ	онтологія	інтерактивний документ	документ
1	Рекурсивний редуктор	10	10	10		10	14	Cibola/Oleada	10	0	0		0
2	General Architecture for Text Engineering	10	0	0		0	15	rumorphy2	10	0	0		0
3	Affx Grammars Over a Finite Lattice (AGFL)	10	0	0		0	16	Russian Morphological Dictionary	10	0	0		0
4	RapidMiner	10	0	0		0	17	Stemka	10	0	0		0
5	LexicoSyntactic Pattern Language (LSPL)	10	0	0		0	18	MonoConc	10	0	0		0
6	Томіта-парсер	10	0	0		0	19	TextAnalyst 2.0	10	0	0		0
7	Конспект (TextTermin)	10	0	0		0	20	AskNet	10	10	0		0
8	ИСИДА-Т	10	0	0		0	21	Vivisimo	10	10	0		0
9	Link Grammar Parser	10	0	0		0	22	Quintura Searchcrystal	10	10	0		0
10	Link Grammar Parser for Russian	10	0	0		0	23	МедіаЛінгва	10	10	0		0
11	RCO (Russian Context Optimizer)	10	0	0		0	24	ABBYY Retrieval & Morphology (ARM) Engine	10	10	0		0
12	libmorphukr	10	0	0		0	25	Apache Lucene/Solr	10	10	0		0
13	Mystem	10	0	0		0	26	Convera RetrievalWare	10	0	10		10

Рисунок 1.7 – Результати оцінки за категорією "Представлення обробленого тексту"



Після проведеної оцінки систем, необхідно визначити ступінь важливості кожного критерію по лінгвістичній шкалі [82]. З огляду на поставлену задачу, критерії розподілено наступним чином:

**Мала важливість** – клас системи;

**Слабка важливість** – вхідні формати файлів xls,xlsx,txt; вихідні формати файлів PDF, JSON, TXT;

**Середня важливість** – підтримка російської мови; підтримка англійської мови; всі доступні види аналізу;

**Значна важливість** – вхідні формати файлів doc/docx і pdf; вихідні формати файлів HTML, XML; виділення з текстів термінів, зв'язків, атрибутів, контекстів, зображень;

**Абсолютна важливість** – підтримка української мови; представлення обробленого тексту у вигляді тексту, веб-документу, онтології, інтерактивного документу.

Результати ранжування представлено на рис. 1.8.

№	Система	Значення	№	Система	Значення
1	Рекурсивний редуктор	0,922	14	Convera RetrievalWare	0,345
2	Apache Lucene/Solr	0,415	15	Link Grammar Parser	0,323
3	AskNet	0,415	16	Link Grammar Parser for Russian	0,323
4	Конспект (TextTermin)	0,385	17	ИСИДА-Т	0,323
5	Томіта-парсер	0,376	18	RCO (Russian Context Optimizer)	0,297
6	Affx Grammars Over a Finite Lattice (AGFL)	0,376	19	MonoConc	0,259
7	General Architecture for Text Engineering	0,376	20	TextAnalyst 2.0	0,259
8	LexicoSyntactic Pattern Language (LSPL)	0,376	21	Mystem	0,234
9	RapidMiner	0,376	22	Cibola/Oleada	0,209
10	Quintura Searchcrystal	0,361	23	libmorphukr	0,209
11	Vivisimo	0,361	24	рymorphу2	0,209
12	ABBYY Retrieval & Morphology (ARM) Engine	0,361	25	Russian Morphological Dictionary	0,209
13	Медіалінгва	0,361	26	Stemka	0,209

Рисунок 1.8 – Результати ранжування систем

Як показують результати оцінювання, система «Поліедр-Рекурсивний редуктор» має значно кращі результати, ніж інші розглянуті програмні засоби. Це пов'язано з тим, що вона об'єднує підтримку релевантних мов, широкі можливості

по аналізу текстів (включаючи морфологічний і синтаксичний аналіз) і засоби для представлення результатів аналізу в різній формі.

#### 1.4. Висновки за розділом 1

В першому розділі проаналізовано літературу за темою дисертаційного дослідження та проведено огляд існуючих засобів для обробки та представлення природномовних текстів. Розроблено критерії оцінки обраних засобів і здійснено їх ранжування. Даний аналіз показав нестачу інструментальних засобів, здатних представляти результати обробки неструктурованих текстових масивів у вигляді, що задовольняв би критерію трансдисциплінарності. Дана ситуація дозволяє поставити такі задачі дослідження:

- 4) розробити концептуальну онтологічну модель і засоби трансдисциплінарного представлення електронних образів видань;
- 5) розробити методи трансдисциплінарного представлення видань;
- 6) створити мережеві програмно-інформаційні засоби трансдисциплінарного представлення інформації, що міститься в електронних виданнях;
- 7) експериментально підтвердити ефективність роботи розроблених моделей і програмно-інформаційних засобів.

Найкращим засобом для трансдисциплінарного представлення інформації в ході аналізу виявлено систему «Поліедр-Рекурсивний редуктор». Висунуто гіпотезу про можливість її розвинення шляхом доповнення класифікатором стилів видань та створенням бази еталонних видань різних стилів для ефективнішої обробки різномірної інформації.

Результати досліджень першого розділу наведено в публікаціях [10, 83, 84].

## РОЗДІЛ 2. МОДЕЛЬ ТРАНСДИСЦИПЛІНАРНОГО ПРЕДСТАВЛЕННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ РІЗНИХ СТИЛІВ

### 2.1. Загальні поняття

#### 2.1.1. Онтології

Онтологія [30, 47, 74, 78, 85–88] – концептуально-понятійний каркас всіх без виключення наукових теорій. Суб'єктивно її можна представити як певне бачення експерта певної картини світу, яка відображає його операціональне середовище. Це бачення представимо певним нарративом свідомості, яке агрегує усі його знання. Однією з ознак онтології нарративу свідомості є формат таксономічного представлення змісту цього нарративу. Таксономія утворює операціональні можливості розуміння нарративу через структури та значення концептів, що його складають, для автоматичної обробки інформації, яка відображається у свідомості експерта.

Професійна діяльність людини, у певному операціональному середовищі, може бути представлена у системному вигляді наступного формату {дія → результати}. Це натуральний тип систем –  $SN$  [89]. Такі системи існують за умови наявності не пустої множини можливих наборів дій –  $F$  [89, 90], яка являє собою набір функцій інтерпретації, що коректно реалізуються в операціональному середовищі, в якому реалізує свої знання та навички людина. Безпосередньо ці функції-інтерпретації задані над об'єктами певної предметної галузі (ПГ), нарративний формат представлення якої задається композиціями визначених концептів.

Концепти представимо у форматі множини  $X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$ . Кожен концепт, у свою чергу, представимо множиною властивостей  $R$ , що має формат декартових добутоків множини  $X$  самої на себе –  $R = \prod_i^n X_i$ .

Множина дій  $F$  також утворюється декартовим добутком множин  $X$  і  $R$ :  $F = X \times R$ . [89, 90].

Згідно [89, 90] існує множина дій  $F_k \subset F$  таких, що завжди існує хоча б одне непусте  $f^i \in F_k$  таке, що завжди знайдеться така множина концептів  $X_j$ , для яких  $f^i(x_1, x_2, \dots, x_n) \in F_k$ . Тобто, справедливе твердження, що для довільного концепту  $X$  завжди знайдеться відповідний набір дій з множини  $F$ .

Згідно [89] такі множини можуть бути замкнуті або відкриті. Але усі множини дій, які реалізуються при умові зв'язності наративу свідомості людини та наративних описів її операціональної реалізації у зовнішній предметній екосистемі, можуть бути завжди доповнені додатковими діями. Тобто вони носять відкритий характер.

Більш того, кожна множина концептів  $X$ , над якою задано набір дій  $F$ , може бути доповнена також новим (новими) концептом типу  $x_{n+1}$ , такими, що існує додаткова властивість  $r'$ , яка забезпечує виконання правила  $f^i(x_1, \dots, x_n, x_{n+1}) \in F_k$ . Таким чином, відкрита множина  $F$  має властивість емерджентності [86, 91], що утворює технологічні умови консолідації знань, що акумулюються у наративі свідомості людини та наративних описах, які характеризують семантику її операціональної реалізації у професійній сфері.

Зазначимо, що дії, що утворюють множину  $F_t \subset F$ , можуть бути представлені у вигляді певних тверджень, які визначаються як у наративі свідомості, так і у наративних описах професійної сфери, що й визначає їх консолідованість. Така консолідованість на рівні послідовностей простих висловлювань й визначає рівні професійного розвитку людини в операціональному середовищі конкретної предметної галузі.

Однак, усі ці дії утворюють натуральну систему. А також ці дії інтерпретують консолідовану функціональність конкретних концептів із наративів свідомості та операціональності. Також кожна така дія, як певна функція, характеризується хоча б однією властивістю. Тоді ми можемо на основі натуральної системи визначити певну онтологію, яка носить трансдисциплінарний характер [74, 80, 89, 92, 93], в

якій множини концептів та концептуальних відношень максимально повні, а до функцій інтерпретації додаються аксіоми:

$$O = \langle X, R, F, A, (D, R_s) \rangle, \quad (2.1)$$

де  $X$  – множина концептів;

$X = \{X_1, X_2, \dots, X_i, \dots, X_n\}, i = \overline{1, n}, n = \text{Card}(X)$  – кінцева множина концептів (понять) заданої ПГ;

$R = \{R_1, R_2, \dots, R_k, \dots, R_m\}, R \subseteq X_1 \times X_2 \times \dots \times X_n, k = \overline{1, m}, m = \text{Card}(R)$  – кінцева множина відношень, які задні над концептами ПГ й визначають їх множинну семантику. Дані відношення визначають тип взаємодії між концептами. Зазвичай відношення поділяють на загальнозначущі (з яких виділяють відношення часткового порядку) і конкретні відношення окремої ПГ;

$F : X \times R$  – кінцева множина функцій інтерпретації (дій), заданих на концептах і/або відношеннях. Окремим випадком визначення множини функцій інтерпретації  $F$  є глосарій, складений для множини понять  $X$ . Визначення поняття  $X_i$ , в загальному випадку, включає підмножина понять  $\{x_{i-1}\}$ , через які визначаються  $X_i$ , відношення, що зв'язує  $X_i$  з  $\{x_{i-1}\}$ , і множина атрибутів (ознак), властивих  $X_i$ ;

$A$  – кінцева множина аксіом, які реалізують формування завжди істинних висловлювань (визначень і обмежень) в термінах операціональної тематики ПГ;

$D$  – кінцева множина додаткових визначень понять в термінах операціональної тематики ПГ;

$R_s$  – кінцева множина обмежень, що визначають область дії понятійних структур в операціональному середовищі людини.

Таким чином ми отримали пару  $(N, O)$ , яка представляє функціональність та структуру консолідованих наративів свідомості та операціональності.

Трансдисциплінарна онтологія натуральної системи є формальним представленням концептуальних знань людини про предметні галузі і може бути

представлена у форматі інформаційної системи. Процес побудови такої ІС являє собою створення композиції певних суджень, тверджень, висловлювань або термінів-понять, а також відношень між ними. Результат даного процесу є основою для побудови однієї зі складових частин наукової теорії – онтологічної бази знань у заданій предметній галузі, описаній в декларативній формі [80, 89]. У зв'язці з натуральною системою, що її утворює, вони реалізують консолідацію наративів свідомості та операціональності людини.

Особливістю трансдисциплінарних онтологій [80, 92, 93], є наявність в їх натуральній системі функцій вибору  $F_{set}$  виду [80, 90, 93, 94]:

$$(f : X \rightarrow \Psi \cong T) \Rightarrow f(T) \subset F_{set}, \quad (2.2)$$

де  $f$  – певна дія;

$X$  – консолідована множина концептів наративів свідомості та операціональності;

$\Psi, T$  – таксономічні основи онтології;

Знак  $\Rightarrow$  визначає, що із виразу зліва витікає вираз, що знаходиться праворуч.

При чому тут функція вибору  $F_{set}$  реалізує визначення концептів та їх контекстів, що консолідують наративи свідомості та операціональності конкретного фахівця у процесі його професійної діяльності.

Однак, в реальності, з довільного наративу не завжди можна виділити усе різноманіття інформації, тому можуть мати місце граничні випадки виразу (2.1) при виконанні умов упорядкування концептів множини  $X$ , виявлення між ними бінарних ієрархій, та інтерпретування відношень, що утворюють ці ієрархії. Всі можливі комбінації упорядкування наративу утворюють різні варіанти онтологічних конструкцій, від простого словника до формального відображення концептуальної бази знань. Зокрема, згідно [74], можна виділити:

- 1) Множина  $X$  – включає концепти, що строго упорядковані послідовностями контекстів, які змістовно визначають певний наратив та документ, тобто це

лінійно-упорядкований (неструктурований) текст  $T^T$ , такі наративи та документи являють собою неупорядковану множину концептів, над контекстами яких задано тільки одне відношення строгого порядку, яке інтерпретується функцією послідовного переходу від одного контексту до іншого;

2) Концепти множини  $X$  упорядковуються лексикографічно, у такому випадку ми маємо документ у форматі глосарію  $O^1$ ;

3)  $X$  розбивається на певні ієрархічні класи на основі бінарних відношень – таксономія  $O^2$ ;

4)  $X$  можна розбити на певні класи, які характеризуються як відношеннями між концептами  $R_t$  так і відношеннями між контекстами, що описують ці концепти  $R^+$ , однак над множиною  $X$  не визначено жодної функції, формально це можна представити у наступному вигляді:  $X \neq \emptyset, (R \neq \emptyset | R = R_t \cup R^+), F = \emptyset$  –; у такому випадку ми маємо формат тезаурусу;

5) Концепти наративу чи іншого документу представлено у форматі онтології. Над концептами множини  $X$  визначено тільки одну інтерпретуючу функцію, тобто маємо наступні умови утворення онтології:  $X \neq \emptyset, R = \emptyset, card(F) = 1$ ; у такому випадку таку онтологію будемо називати простою –  $O$ ;

6) Маємо множину концептів  $X$ , множину відношень між ними, а також певний набір інтерпретуючих функцій:  $X \neq \emptyset, R \neq \emptyset, card(F) > 1$ , у такому випадку будемо визначати таку онтологію повною –  $O_t$ .

Така класифікація онтологій за функціональними ознаками відповідає описам, наведеним в роботах [50, 51, 100–109, 78, 82, 88, 95–99].

### 2.1.2. Таксономічні основи структуризації документів

Кожна онтологія має в своїй основі певну множину таксономій. На їх основі можуть формуватися певні класифікатори, які відображають семантику та

взаємозв'язки концептів нарративних документів. Технологічним інструментом для формування та супроводу класифікаторів є таксономія [80, 110], що відображає певну ієрархію взаємодії концептів нарративних документів.

Ієрархія у таксономічних системах задається за допомогою бінарних відношень, що визначають характер взаємодії між концептами мережевих документів, що обробляються. На підставі зазначених відношень, що визначають взаємодію між концептами, реалізується процедура розбиття множини концептів на класи [80, 93, 111]. Для побудови предметних класифікаторів використовуються таксономічні категорії. Список зазначених категорій визначається в процесі розв'язання прикладної задачі на основі властивостей, які об'єднують концепти за тематичними ознаками. Складність множини взаємовідношень між ними можна побачити на прикладі часто використовуваної наступної таксономічної структури:

$\tilde{T} = \{тип; підтип; клас; підклас; ряд; підряд; родина; підродина; рід; підрід; вид; підвид; різновидність; форма\}$ .

Для формування будь-якої таксономічної категорії може бути використано бінарне відношення – «бути елементом категорії». Поняття *категорія* тут відіграє роль змінної і може приймати будь-яке значення з вищенаведеного списку. Також кожна категорія є складним концептом, якому, як певному поняттю предметної галузі, відповідає конкретне ім'я. Дане відношення дозволяє організувати ієрархію концептів на рівні визначення їх приналежності до конкретного тематичного класу.

Тут відіграє істотну роль поняття складності концепту. Мається на увазі, що концепт має просту структуру (простий концепт), якщо структурно він не може бути розкладений на інші концепти. Тоді поняття складного концепту може бути рекурсивно визначено як поняття, що структурно включає непорожню множину простих концептів. Тобто простий концепт пов'язаний зі складним відношенням – «бути елементом категорії».

Проте слід враховувати, що концепти мережевих документів представлені тільки конкретними іменами. Їх активне використання в операціональному



просторі експерта вимагає визначення їх смислів, які дають можливість інтерпретувати конкретні концепти при виборі дій, застосовних до них у процесі вирішення предметних задач. Такі смисли можуть бути задані як бінарними відношеннями між активними концептами, так і унарними властивостями самих концептів, що характеризують їх прикладні ознаки.

У загальному вигляді бінарні відношення таксономічних категорій можуть бути представлені простою моделлю бінарних відношень «об'єкт – об'єкт» [80, 110–112]. Просту модель бінарних відношень можна узагальнити до рівня множинних відношень за типом «група об'єктів – об'єкт». Тоді відношення між таксономічними категоріями можна розглядати як деякі гіпервідношення  $Gr$  виду –  $YGrx$ , де  $Y$  – множина всіх можливих множин концептів  $X$  таксономічної категорії  $\tilde{T}$ , а  $x$  – один з концептів цієї множини. Зрозуміло, що, якщо розглядати множинне бінарне відношення типу «група об'єктів – група об'єктів», то його можна представити у вигляді певної множини гіпервідношень  $\{Gr\}$ , що може бути відображено через множину бінарних відношень за типом «група об'єктів – об'єкт». Тому коректним є надалі розглядати саме гіпервідношення  $Gr$ , які завжди можна визначити до множини групових бінарних відношень, взявши до уваги їх декартовий простір [113].

Тобто для гіпервідношення  $Gr$  можна сформувати множину, елементи якої забезпечують нумерацію усіх концептів таксономії, а їх упорядкування за правилом монотонного зростання над множиною натуральних чисел  $N$ , підмножиною якої є номери концептів. Елементи цієї множини, яку складає кінцева послідовність натуральних чисел, забезпечують визначення порядку взаємодії усіх концептів таксономії, основою якого є бінарні взаємодії сусідніх концептів, які визначають мережеві документи, що обробляються. Визначена гіпервпорядкованість представима у форматі множин бінарних впорядкованостей  $P$  і представлена у вигляді наступного виразу –  $P = \{p_n\} n \in N$ :

$$xy \Leftrightarrow \forall n \in N : xp_n y \quad (2.3)$$

Множину гіпервідношень  $Gr$ , характеризують певні властивості, які реалізують бінарні відношення як між концептами, так і між концептами та таксономіями, а також між таксономіями [90, 112, 113]:

– агіперциклічність – якщо для  $Gr$  не існує гіперциклічної множини  $X \subseteq U$  такої, коли:

$$\forall x \in X \exists Y \subseteq X : YGrx \quad (2.4)$$

– іррефлексивність:

$$YGrx \Rightarrow (Y / \{x\})Grx \quad (2.5)$$

– гіпертранзитивність:

$$YGrx, x \in X, XGrz \Rightarrow ((Y \cup X) / \{x\})Grz \quad (2.6)$$

– регулярність:

$$YGrx, Y' \supseteq Y \Rightarrow Y'Grx \quad (2.7)$$

Множинна впорядкованість, як властивість таксономічних категорій, забезпечує реалізацію процедури додавання/видалення множин концептів таксономій мережевих документів та їх підмножин концептів, пов'язаних між собою конкретним бінарним відношенням впорядкованості. Таким чином, якщо множина гіпервідношень володіє набором властивостей (2.3) – (2.7), забезпечує відображення різних активних станів взаємодії із мережевими документами на основі виділення бінарних відношень часткової упорядкованості  $\tilde{p}$  для множин концептів  $\{X\}$ , то довільна таксономія  $\tilde{T}$  може бути представлена предикативними виразами з концептів, які завжди приймають значення істинності. Конструктивність цього твердження полягає у тому, що у множині концептів  $X$  завжди існує непуста підмножина концептів  $\{x\}$ , таких що пов'язані між собою бінарним відношенням часткової впорядкованості  $\tilde{p}$ .

$$YGrx \Rightarrow \exists y \in Y : y\tilde{p}x \quad (2.8)$$

Тобто, відношення часткової впорядкованості  $\tilde{p}$ , може бути виділено із множини гіпервідношень  $Gr$ , з одного боку, а з іншого боку, відношення  $\tilde{p}$ , може бути індуктивно інкапсульоване у множину  $Gr$ . Тоді, на її основі, у довільної таксономії  $\tilde{T}$  над концептами визначається множинне відношення – «група об'єктів – об'єкт», що має властивості (2.4) – (2.8).

$$YGrx \Leftrightarrow n \in N \exists y \in Y : y p_n x : p_n \in P \subseteq Gr \quad (2.9)$$

Відзначимо, що предикативні вирази, що формулюються на основі концептів таксономії з заданим множинним відношенням упорядкованості виду (2.1) – (2.7) приймають лише значення істинності. Це дозволяє формувати на основі термінів-концептів таксономічної системи мережевих документів лінгвістичні вирази, що відображають семантичні стани мережевих ресурсів.

Таким чином довільну таксономію можна представити у форматі орієнтованого графу без циклів, що утворюється концептами, які ієрархічно взаємозв'язані між собою. Вона може бути представлена у вигляді упорядкованої пари виду:

$$T = \langle X, R_t \rangle, \quad (2.10)$$

де  $T$  – таксономія,

$X$  – множина концептів,

$R_t$  – множина бінарних відношень між концептами таксономії  $T$ .

Категорія таксономій забезпечує врахування семантики наративів [114–116], що надає їй певні напрями застосування:

- таксономії визначають концептуальну структуру онтології [115–117];
- вузли таксономій містять контекстні описи їх сенсів, що дозволяє вважати їх концептами;
- таксономії утворюються з класів концептів, що характеризує їх певними властивостями згідно визначення онтології (2.1);

– над таксономіями можна визначати множинні відношення, які їх зв'язують між собою, при чому ці відношення є бінарними, які встановлюють відношення між певними контекстами, що описують чи визначають концепти онтології.

Таксономія (2.10) характеризується відношеннями між концептами наративу й може бути представлена орієнтованим графом без циклів [34, 115, 116].

Як бачимо формула (2.1) може бути отримана із формули (2.10) упорядкованим включенням категорій, що визначають предметну галузь, з додаванням натуральної системи  $SN$ .

Якщо визначити контексти вузлів таксономій як елементи певних знань, тоді їх сукупність відповідає певному фрагменту картини світу. Розгляд всього різноманіття таксономій виявляє гіпервластивість рефлексії, що реалізується як відображення усіх таксономій самих на себе. Виходячи з факту, що з концептів, які утворюють вузли таксономії, формулюються конкретні висловлювання, що мають істинні значення [34, 113, 115], рефлексивне відображення таксономії самої на себе представимо у вигляді спеціальної натуральної системи  $SN_T$ .

Така  $SN_T$  може мати характеристичний предикат виду:

$$Pr(x_1, \dots, x_n) = 0 \Rightarrow \exists T \subseteq O : \forall x \in X \exists SN_T \subseteq O : T = YG_x, \quad (2.11)$$

де  $x_n$  – прості концепти таксономії  $T$ ;

$G$  – гіпервластивість, яка утворює певний непустий клас концептів таксономії, у даному випадку –  $Y$ .

Згідно [115] рефлексія визначає  $SN_T$ . Позначимо рефлексію через  $F_f$ .

$$(F_f : X \rightarrow X \rightarrow Y) \quad (2.12)$$

Відображення (2.12) представимо у рекурсивному вигляді:

$$F_f = \begin{cases} F_f(X, SN_T) \rightarrow SN_T \\ T \\ XG_x \end{cases} \quad (2.13)$$

Правила (2.10) – (2.13) описують процес побудови натуральної системи  $SN_T$ , яка визначає усі функції – інтерпретації відношень між концептами таксономії. Додавання натуральної системи дозволяє нам побудувати онтологію вибраних наративів.

### 2.1.3. Зростаюча пірамідальна мережа

Одним із конструктивних форматів формального представлення таксономій є зростаючі пірамідальні мережі (ЗПМ) [118]. ЗПМ являє собою ациклічний орієнтований граф, в якому немає вершин, що мають одну вхідну дугу. Вершини, які не мають вхідних дуг, називаються рецепторами (або терміналами), інші – концепторами (або класами). Підграф пірамідальної мережі, що містить вершину  $a$  і всі вершини, від яких є шляхи до  $a$ , називається пірамідою вершини  $a$ . Вершини, що входять в піраміду вершини  $a$ , утворюють її субмножину. Множина вершин, до яких є шляхи від вершини  $a$ , називається її супермножиною.

За означенням [86, 119–121] зростаючою пірамідальною мережею  $Q$  називається ациклічний орієнтований граф  $Q = (U, E)$ , в якому відсутні вершини, які мають одну дугу, що заходить.  $U = \{u_i | i = \overline{1, n_1}\}$  – множина вершин мережі, де  $u_i$  – довільна вершина мережі,  $n_1$  – кількість вершин у мережі.  $E = \{e_i | i = \overline{1, n_2}\}$  – множина дуг мережі, де  $e_i$  – довільна дуга,  $n_2$  – кількість дуг у мережі. В якості прикладу пірамідальної мережі, наведемо її структуру, що представлено у роботах В.П. Гладуна та В.Ю. Величко [118–120] (рис. 2.1, 2.2)

Згідно класичного підходу до формування ЗПМ [118–121], вони мають два види вузлів: рецептори та концептори. Рецептори визначаються конкретними значеннями ознак, які присутні в описах концептів. Концептори формуються на основі комбінацій значень ознак, які утворюють властивості, що дозволяє ідентифікувати концепт. Ці комбінації також можуть визначатися спільними компонентами описів декількох концептів. Концептори можуть мати ребра, що до

них заходять. Тоді концептор взаємодіє з іншими концепторами або рецепторами ребрами, які можуть мати ім'я.

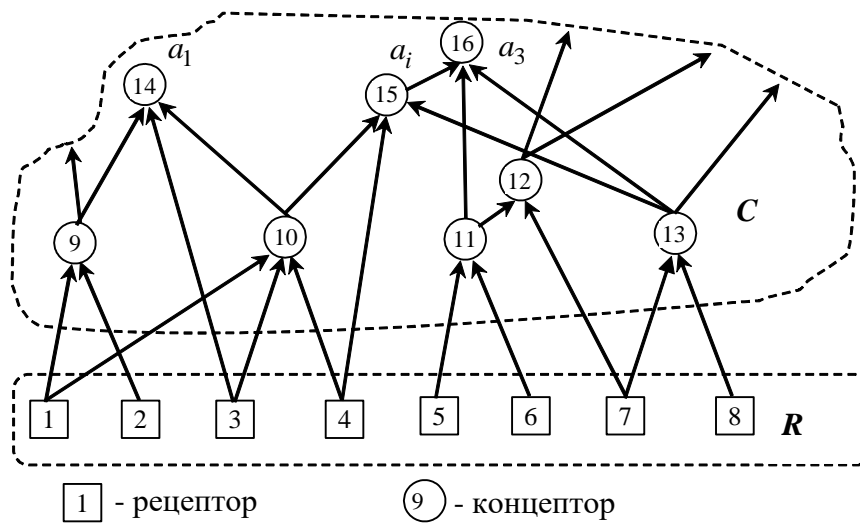


Рисунок 2.1 – Приклад семантичної мережі пірамідальної структури

На рис. 2.2 зображено фрагмент процесу формування ЗПМ. Певний стан ЗПМ зображено на рис. 2.2 а. Після додавання до мережі нового концепта  $a_i$  в мережі з'являється новий додатковий концептор з номером 10 -  $sc_{10}^3$  (рис. 2.2 б), який визначає факт зв'язності концептора 14 і концептора 15.

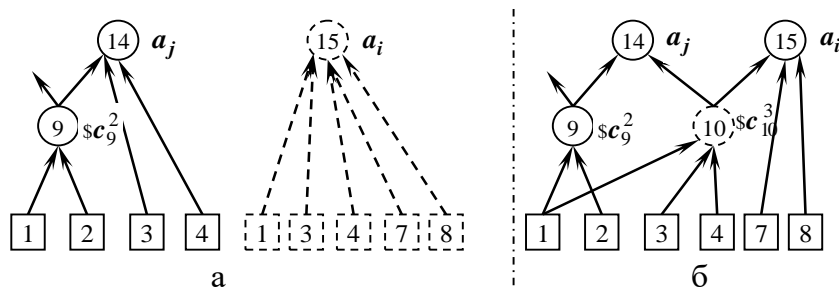


Рисунок 2.2 – Фрагмент формування пірамідальної мережі

ЗПМ має властивість додавати новітні концептори мережі, які утворюються на основі спільних ознак концепторів, які утворюють стани взаємодії. Імена таких концепторів, що додаються до мережі в процесі формування ЗПМ, позначаються порядковими номерами з додаванням попереду номера символу \$. У процесі додавання нових концептів до мережі вони утворюють вузли ЗПМ, й визначають відношення взаємодії між собою.

Вказана властивість ЗПМ досить конструктивна для формування таксономій (таксономічних моделей) наративів і документів. Кожна таксономічна модель у форматі ЗПМ реалізується на основі представлення вхідної інформації, яка специфікує процес, що досліджується у форматі текстового файлу. Маркування вказаного файлу повинно включати наступне: рядок заголовка, рядки з описами властивостей об'єктів. Структура рядка заголовка: об'єкт, клас, (ім'я ознаки 1), ..., (ім'я ознаки n). Імена «об'єкт» і «клас» є обов'язковими в рядку заголовка. Розділовий знак між елементами рядка – кома. Структура таких рядків:

(Ім'я об'єкта 1), (ім'я класу), (значення ознаки 1), ..., (значення ознаки n)

(Ім'я об'єкта m), (ім'я класу), (значення ознаки 1), ..., (значення ознаки n).

Різні об'єкти можуть мати різну кількість ознак, які наводяться в описах об'єктів у довільному порядку. Формат вхідного файлу без рядка заголовка застосовується для побудови логіко-лінгвістичних моделей семантичного аналізу текстових документів, де об'єкти описуються довільною кількістю різноманітних ознак. Зв'язки між об'єктами встановлюються на основі збігу імен ознак в описах різних об'єктів.

Як можна визначитися із вищенаведеного, формат ЗПМ реалізує теоретико-графовий підхід до таксономізації. На основі ЗПМ реалізуються графові алгоритми, та формуються і відображаються ієрархії між об'єктами що складають пірамідальну мережу. Також до характерних рис ЗПМ при таксономізації слід відзначити той факт, що вони реалізують як агломераційні (об'єднувальні) так і дивізійні (розділяючі) алгоритми [89]. Більш того, як можна побачити з рис. 2.1 та рис. 2.2, ми можемо виділяти із ЗПМ певні таксономічні структури [91], які відображають ієрархії переходу між різними операціональними станами діяльності експерта.

Своєю чергою, таксономії як певний формат відображення структури наративу, досить об'єктивно представляють структуру довільного операціонального середовища, яке формується певними функціями дії, що породжуються онтологіями.

У субмножині та супермножині вершини виділяються 0-субмножина та 0-супермножина, що складаються з тих вершин, які пов'язані з нею безпосередньо. При побудові мережі вхідною інформацією служать набори значень ознак, що описують деякі об'єкти (атрибути). Рецептори відповідають значенням ознак. У різних завданнях це можуть бути об'єкти, класи об'єктів, імена властивостей, відносини, стани, дії. Для опису функціональних стилів це – способи розмітки документів. Концептори відповідають описам об'єктів в цілому і перетину описів.

Також ЗПМ можна представити у вигляді множини навантажених дводольних графів  $G$  [49].

$$G = (V1 \cup V2, E), \quad (2.14)$$

де  $V1 \cap V2 = \emptyset$ , вершини з  $V1$ , розмічені іменами предикатів, а вершини з  $V2$  – іменами аргументів;  $E$  – множина дуг (ребер). Дуги графа з'єднують вершини, помічені іменами предикатів, з вершинами, поміченими іменами аргументів.

Вершини з множини  $V1$  називаються вузлами-предикатами, вершини з множини  $V2$  – вузлами-концептами, а самі предикати – концептуальними предикатами.

Висловлювання формується на основі композиції вершин, інцидентних до одного ребра.

Алгоритм формування ЗПМ подається згідно робіт [49, 86]:

1. Визначити першу вершину (праву або ліву) за напрямком відношення, при умові, що воно не комутативне.
2. Обрати праву/ліву вершину та інцидентне ребро.
3. Обирати праву/ліву вершину з інцидентним ребром, яке має ліву/праву вершину.
4. Дводольний граф визначити як висловлювання.

Після цього необхідно обчислити значення висловлювання: якщо воно істинне – вершини включаються до множини об'єктів інтерфейсу, якщо хибне – вершини не входять до цієї множини.



Як бачимо, даний алгоритм забезпечує побудову множини дводольних графів, які мають спільні вершини, у вигляді зв'язного графу, що не має циклів.

#### 2.1.4. Метод рекурсивної редукції

Операціональність довільного наративу [122] виду (2.10) визначається крім натуральної системи ще гіпервластивостями рефлексії –  $R_f$ , рекурсії –  $R_k$  і редукції –  $R_d$ , які визначають замкнуту множину (2.15), й забезпечують зв'язність та динамічне переупорядкування концептів та їх контекстів у наративі [80, 123]. Зазначені гіпервластивості володіють трансдисциплінарним та інтерпретаційним характером [124]:

$$\mathfrak{R3} = \{R_f, R_k, R_d\}, \quad (2.15)$$

де: рефлексія – забезпечує реалізацію консолідованої системної зв'язності у масивах Big Data, мінімізацію змістовних розривів між контекстами, моніторинг щодо стійкості їх поведінки під час активації; та модифікації їх таксономій;

рекурсія – забезпечує кінцеву циклічність використання концептів онтології, які можуть визначати певні процеси, явища і форми при виявленні між ними певних взаємозв'язків при умові їх активації в операціональному середовищі їх використання;

редукція – реалізує декомпозицію складних концептів і об'єктів [76].

Когнітивність зазначених гіпервластивостей виражається в можливості інтерпретації їх гіперфункцій, таких як: аналіз і синтез Big Data, їх структуризація, вибір і прийняття рішень, які, відповідно до [89, 122, 124, 125], визначаються як когнітивні.

Поняття «онтологія», «нاراتив», «рефлексія», «рекурсія», «редукція» визначають категорію трансдисциплінарності. Розглядаючи когнітивно-комунікативний процес взаємодії з консолідованими IP, під трансдисциплінарністю системи знань будемо розуміти умови застосування рефлексивно-активної рекурсивної редукції  $\mathfrak{R3}$  [109]. А це означає, що (2.42) онтологію виду (2.2) можна розширити множиною  $\mathfrak{R3}$  і гіперфункцією її інтерпретації.

Метод рекурсивної редукції [109] призначений для структуризації ПМ текстів, та представлення їх у вигляді онтології. Перетворення структуризації тексту є багатоетапним процесом, який можна представити виразом:

$$T^T \rightarrow T_{SN} \rightarrow O^1 \rightarrow O^2 \rightarrow O \quad (2.16)$$

де  $T^T$  – ПМ текст;

$T_{SN}$  – первинна структура тексту;

$O^1$  – глосарій;

$O^2$  – таксономія;

$O$  – онтологія.

На першому етапі перетворення ПМ тексту на первинну структуру тексту,  $T^T \rightarrow T_{SN}$  виконується його лексичний аналіз, що потребує використання значної кількості лінгвістичної інформації. Даний процес виконується окремою системою – лексичним аналізатором. Подальший процес  $T_{SN} \rightarrow O$  передбачає послідовне перетворення первинної структури тексту на онтологію шляхом застосування певного набору динамічно заданих користувачем правил.

Отже, до вхідного ПМ тексту застосовується спеціалізований оператор (2.17), що рекурсивно виконує процес його редукції.

$$F_{rd} : T_{SN} \rightarrow O \quad (2.17)$$

Оператор редукції описується в термінах  $\lambda$ -теорії і являє собою комбінацію чотирьох операторів. При цьому три з операторів виконують кроки перетворення задані виразом  $T_{SN} \rightarrow O^1 \rightarrow O^2 \rightarrow O$ , а один виконує допоміжну функцію:

$$F_{rd} = F_{l^*} \circ F_x \circ F_{smr} \circ F_{ct}, \quad (2.18)$$

де  $F_{l^*}$  – оператор агрегації, що виконує допоміжну функцію, перетворюючи множину лексем в множину конструктів;

$F_x$  – оператор ідентифікації концептів;

$F_{smr}$  – оператор ідентифікації онтологічних зв'язків;

$F_{ct}$  – оператор ідентифікації контекстів.

Кожен з цих операторів реалізує один крок перетворення. Повний його цикл реалізує утворення структури певних контекстів, які містяться у вхідному тексті, після чого перетворення рекурсивно викликається знову, до повного виділення всієї інформації.

В загальному випадку оператор виконання перетворення  $F$  задається базою правил  $G_R$  виконання даного перетворення. Правило  $g \in G_R$  має уніфіковану для всіх етапів структуру:

$$g = f_{ap}^g, f_{tr}^g \quad (2.19)$$

де  $f_{ap}^g$  – функція застосовності, яка визначає умови застосування даного правила до певного набору вхідної інформації;

$f_{tr}^g$  – функція перетворення, що задає перетворення вхідної інформації.

Кожна функція застосовності  $f_{ap}^g$  – це лямбда-термін [126] форми:

$$f_{ap} = (\lambda l_1, \dots, l_n. t_l(\bar{l}) \& t_r(\bar{l})) x_1, \dots, x_n \quad (2.20)$$

де  $l_i$  – змінна, яка приймає значення на множині лексем;

$x_i$  – аргумент функції, що визначає значення  $l_i$ ;

$t_l$  – функція ідентифікації лексеми;

$t_r$  – функція ідентифікації відношення.

Функція ідентифікації лексеми визначається набором предикатів ідентифікації лексеми та має вигляд (2.21). Предикати призначені для перевірки різних аспектів певної лексеми з вхідної послідовності.

Основним типом таких предикатів є предикати перевірки ключових слів (2.22). Інші стандартні типи – це предикат, що перевіряє характеристики (2.23), і нульовий предикат (2.24). Нульовий предикат використовується в ситуації, коли очікуються нерегулярні лексеми. Ці лексеми можуть бути відсутні у словниках ключових слів, та існує ймовірність неправильної ідентифікації їх характеристик під час лексичного аналізу. Також можна використовувати спеціалізовані типи предикатів у межах конкретних завдань, наприклад, предикати регулярних виразів.

$$t_l(l_1 \dots l_n) = c_1(l_1) \& c_2(l_2) \& \dots c_n(l_n), \quad (2.21)$$

де  $c_i$  – предикати ідентифікації лексеми.

$$c(l) = \begin{cases} 1, l^T \in L_c \\ 0, l^T \notin L_c \end{cases}, \quad (2.22)$$

де  $l^T$  – текстове подання лексеми  $l$ ;

$L_c$  – словник ключових слів, перевірений предикатом  $c$ .

$$c(l) = \begin{cases} 1, p_c \in P_l \\ 0, p_c \notin P_l \end{cases}, \quad (2.23)$$

де  $P_l$  – сукупність характеристик лексеми  $l$ ;

$p_c$  – характеристика, перевірена предикатом  $c$ .

$$c(l) = 1 \quad (2.24)$$

Функція ідентифікації відношень має вигляд (2.25) і визначається набором предикатів ідентифікації відношень. Предикати мають форму (2.26) і визначаються типом зв'язку, який повинен існувати між певними лексемами у вхідній послідовності. Загалом, не всі лексеми пов'язані з відношеннями, тому більшість предикатів є нульовими предикатами (2.27).

$$t_r(l_1 \dots l_n) = \begin{cases} 1, \forall i, j \in [1..n], i \neq j, r_{ij}(l_i, l_j) \\ 0, \exists i, j \in [1..n], i \neq j, \neg r_{ij}(l_i, l_j) \end{cases}, \quad (2.25)$$

де  $r_{ij}(l_i, l_j)$  – предикати ідентифікації відношень.

$$r(l_1, l_2) = \begin{cases} 1, \langle l_1, l_2, k_r \rangle \in R_{sn} \\ 0, \langle l_1, l_2, k_r \rangle \notin R_{sn} \end{cases}, \quad (2.26)$$

де  $k_r$  – тип відношення, який перевіряється предикатом

$$r(l_1, l_2) = 1 \quad (2.27)$$

### 2.1.5. Онтологічний інтерактивний документ

Отримання структуризованого представлення тексту у вигляді онтології дає широкі можливості експертам для автоматичної обробки наявної в цьому тексті інформації. Проте, дана форма представлення тексту часто виявляється не зручною для користувача. Онтологічний інтерактивний документ є різновидом інформаційної системи, що використовує онтології як джерело інформації як для відображення кінцевому користувачеві, так і для визначення доступних послуг. Найпростіший інтерактивний онтологічний документ має форму (2.28). Основним елементом інтерактивного документа є натуральна система, яка забезпечує інтерактивний доступ до інформації, що міститься в онтології [80].

$$\langle O, SN \rangle \quad (2.28)$$

Натуральну систему можна представити як функцію (2.29)

$$\tilde{y} = \tilde{f}(\tilde{x}^1 \dots \tilde{x}^n), \quad (2.29)$$

де  $\tilde{x}^i$  – дії, що виконуються користувачем при взаємодії з інформацією;

$\tilde{y}$  – результат роботи системи у вигляді текстової та графічної інформації;

$\tilde{f}$  – цільова функція системи, яку можна представити (2.30).

$$\tilde{f}(\tilde{x}^1 \dots \tilde{x}^n) = D(Q_{n-1}(\dots Q_1(X, \tilde{x}^1) \dots, \tilde{x}^{n-1}), \tilde{x}^n), \quad (2.30)$$

де  $X$  – сукупність об'єктів початкової онтології  $O$ ;

$Q_i$  – функції обробки інформації;

$D$  – функція відображення інформації.

Приклади функцій обробки інформації включають функції ієрархічної (2.31) та атрибутивної (2.32) фільтрації [109].

$$Q_h(X, x^*) = \{\bar{x} \in X \mid \bar{x} R x^*\}, \quad (2.31)$$

де  $X$  – сукупність об'єктів онтології або певна їх підмножина;

$x^*$  – об'єкт, що фільтрується;

$\bar{R}$  – конкретний зв'язок між об'єктами.

$$Q_a(X, A) = \{\bar{x} \in X \mid A \cap A_{\bar{x}} = A\}, \quad (2.32)$$

де  $A$  – атрибути, що фільтруються;

$A_x$  – атрибути об'єкта  $x$ .

Ці функції дозволяють вирішити широкий спектр проблем при відображенні різноманітної інформації, включаючи інформацію, отриману з просторово розподілених джерел, які можуть мати різний формат та структуру.

Для забезпечення найбільш ефективної роботи експертів з інформацією, поданою у формі інтерактивного документа, функції обробки та відображення інформації, доступної в рамках відповідної натуральної системи, повинні враховувати особливості проблеми, які необхідно вирішити. Більше того, у більш складних предметних галузях можуть існувати різні типи даних, які кардинально відрізняються за структурою та вимагають особливих підходів.

Цей процес вимагає наявності ефективних засобів налаштування натуральних систем, які дозволять змінювати їх склад і структуру з мінімальним часом. Такі зміни можуть передбачати:

- Додавання спеціалізованих функцій відображення інформації;
- Додавання спеціалізованих функцій обробки інформації;
- Додавання спеціалізованих функцій попередньої обробки даних;

Спеціалізовані онтології (керуючі онтології) можуть бути використані як механізм визначення конфігурації натуральної системи.

## **2.2. Концептуальна онтолого-керована модель трансдисциплінарного представлення ІР**

Концептуальна онтолого-керована модель трансдисциплінарного представлення ІР є описом технології автоматизованої обробки різностильових документів глобального середовища та представлення їх у формі, придатній для подальшого їх перетворення в активний формат інтерактивних баз знань. Практична реалізація такої технології здійснюється на основі WEB-орієнтованих сервісів Когнітивної інформаційної технології (КІТ) «Поліедр».

Математична модель може бути представлена наступною структурою [10]:

$$M = \langle M_{Idn}, M_{Cls}, M_{Cons} \rangle, \quad (2.33)$$

де  $M_{Idn}$  – модель онтолого-керованої класифікації та ідентифікації інформації;

$M_{Cls}$  – модель класифікації природномовного тексту за стилем;

$M_{Cons}$  – модель трансдисциплінарної консолідації інформаційних ресурсів.

### 2.2.1. Модель онтолого-керованої класифікації та ідентифікації інформації

Для ідентифікації інформації може бути використаний метод рекурсивної редукції. Визначено, що створення правил, які використовуються для задання шаблонів ідентифікації, саме в предикативній формі є суттєвим недоліком для некваліфікованих користувачів. Ефективність процесу створення бази правил можна значно підвищити, побудувавши перетворення, яке приводило б до предикативної форми онтологічно задані описи інформації, яку необхідно ідентифікувати. В такому випадку можна говорити про процес онтолого-керованої рекурсивної редукції.

Онтологія, що задаватиме такий опис, повинна максимально повно описувати ПГ, тобто бути активною онтологією виду (2.1). Однак користувачам зручно редагувати стандартні онтології виду (2.34). Для забезпечення максимальної ефективності процесу створення правил слід інтерпретувати множину атрибутів стандартної онтології як елементи активної онтології (2.1) [83].

$$O_{St} = \langle X, R, F(A'), A(A'), D(A'), R_s(A') \rangle \quad (2.34)$$

Для цього необхідно побудувати перетворення інтерпретації атрибутів:

$$O_s \xrightarrow{G_{O_i}} O_{St} \quad (2.35)$$

Особливості даного перетворення:

- Кінцеві множини функцій інтерпретації  $F$  і додаткових обмежень  $R_s$  створюються з допомогою спеціалізованої процедури;

- Множина додаткових визначень понять  $D$  може бути представлена з допомогою певних атрибутів без додаткової інтерпретації;
- Скінченна множина аксіом  $A$  не використовується в задачі ідентифікації інформації, тому можна прийняти  $A = \emptyset$ .

Активна онтологія виду (2.1) може бути перетворена в набір правил для рекурсивного редуктора за допомогою спеціалізованого перетворення:

$$R_s(A') \xrightarrow{G_{rd}} f_{ap}^g \quad (2.36)$$

$$F(A') \xrightarrow{G_{rd}} f_{tr}^g, \quad (2.37)$$

де  $G_{rd}$  – перетворення формування редуктивних правил.

Складова формування функції застосовності (2.36) є достатньо простою і зводиться до інтерпретації відповідних атрибутів як предикативних виразів, після чого метод рекурсивної редукції може застосовуватися без змін. Складова формування функції перетворення (2.37) потребує модифікації самого методу для того, щоб динамічно враховувати задані онтологією формати результату.

В першу чергу, функція перетворення  $f_{tr}^g$  призначена для формування об'єктів та відношень у результуючій онтології (2.1). Однак в задачі обробки текстів різних стилів додатково потрібна класифікація документів. Для коректної класифікації необхідно крім власне назв об'єктів, виділяти і їх контексти. Для кожного із компонентів оператора редукції (2.18) функції перетворення, що містяться у відповідних правилах, матимуть певну структуру. Однак усі вони засновані на функції створення імен (2.38), яка, у свою чергу, використовує функції форматування текстових подань окремих лексем. Функції форматування можуть бути дуже різними залежно від виконуваного завдання.

$$N(l_1, \dots, l_n) = \sum_{i=1..n} f_i^N(l_i^T), \quad (2.38)$$

де  $f_i^N$  – функція форматування тексту;

$l_i^T$  – подання тексту лексеми  $l_i$ .



Основними операціями, які можна виконувати в результаті форматування, є:

- Нульова операція, тобто використання текстового подання без змін;
- Пропуск маркера, тобто не включаючи його текстове представлення в остаточну назву;
- Нормалізація лексеми, тобто скорочення відповідного слова в називному відмінку однини (для лексем, що представляють слова);
- Нормалізація формату чисел (для лексем, що представляють числа);
- Нормалізація геопросторової інформації – форматування різних типів координат, при необхідності – з перетворенням між системами координат;
- Нормалізація дат – аналіз текстових подань дат та приведення їх до єдиного формату.

Функції перетворення для різних стадій рекурсивної редукції мають таку структуру:

Функція створення об'єкта використовується в правилах, що визначають оператор ідентифікації об'єкта  $F_x$  і має вигляд (2.39)

$$f_{ir}^g(l_1, \dots, l_n) = X(N(l_1, \dots, l_n)), \quad (2.39)$$

де  $X$  – операція створення об'єкта з заданою назвою.

Функція створення зв'язків є більш складною (2.40). Ці функції використовуються в правилах, що визначають оператор ідентифікації відношень. Завдання цієї функції полягає у визначенні двох об'єктів та створенні зв'язку між ними.

$$f_{ir}^g(l_1, \dots, l_n) = R(X(N(l_1, \dots, l_m)), X(N(l_{m+1}, \dots, l_n))), \quad (2.40)$$

де  $X$  – операція створення (або виділення) об'єкта з заданою назвою;

$R$  – операція створення зв'язку між двома заданими об'єктами;

$m$  – індекс, який вказує межу між іменами об'єктів у вхідній послідовності лексеми.

Функція створення атрибутів має вигляд (2.41). Цей тип функції часто вимагає спеціалізованих процедур для визначення об'єкта, до якого належить атрибут. У

загальному випадку це вимагає певного збереження стану обробки тексту (наприклад, запам'ятовування останнього ідентифікованого об'єкта).

$$f_{ir}^s(l_1, \dots, l_n) = A(N(l_1, \dots, l_m), N(l_m, \dots, l_n)), \quad (2.41)$$

де  $A$  – операція створення атрибута за його ім'ям та значенням;

$m$  – індекс, що вказує межу між іменем та значенням.

Результатом роботи модифікованого таким чином методу рекурсивної редукції, буде так звана виколота онтологія, що матиме вид (2.42).

$$O_{mr} = \langle X, R, K(X) \rangle, \quad (2.42)$$

де  $K(X)$  – контексти об'єктів  $X$ , що являють собою особливий випадок атрибутів  $A'$ .

### 2.2.2. Модель класифікації природномовного тексту за стилем

Для вирішення задачі класифікації текстів за стилями може використовуватись механізм ЗПМ [127]. Даний спосіб забезпечує формування понять, ієрархічне упорядкування та класифікацію вхідної інформації на основі виявлення існування об'єднуючої властивості, яка є загальною для двох понять.

Для здійснення класифікації необхідний власне опис стилів – у вигляді онтологічного реєстру (рис. 2.3). Для побудови онтологічного реєстру стилів експертним чином був оброблений масив наукових, навчальних, законодавчих, відомчих, публіцистичних і, частково, художніх документів. В результаті була побудована таксономічна структура функціональних стилів мови, а також визначені атрибути об'єктів, що відносяться до основних стилів.

Такий реєстр може бути представлений як активна онтологія за схемою (2.1), і в подальшому використовуватись в якості бази правил для методу онтолого-керованої рекурсивної редукції.

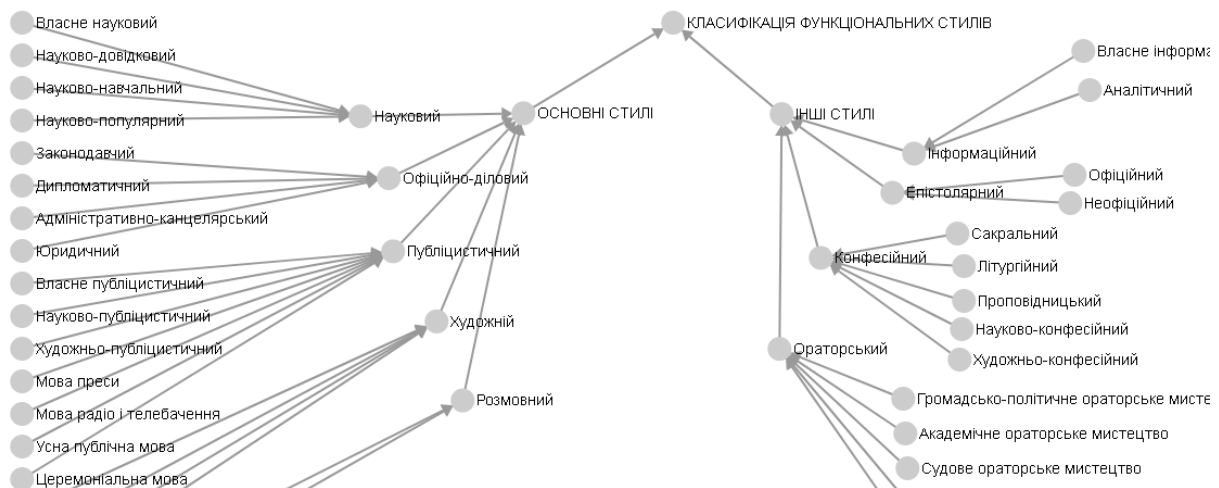


Рисунок 2.3 – Онтологічний реєстр стилів

Об'єктами даного класифікатора є типові види документів, що належать до відповідного стилю. Об'єкти, в свою чергу, мають атрибути, які і описують особливості оформлення тексту конкретних документів.

Додаткові правила формуються на основі об'єктів розробленої онтології з допомогою спеціалізованого перетворення інтерпретації (2.43).

$$G_i : x \rightarrow f_{ap}^g, \quad (2.43)$$

де  $x \in X$  – об'єкт, що належить відповідній онтології, яка представляє класифікатор.

Функція застосовності, сформована на основі такого правила, буде задаватись множиною предикатів ідентифікації, сформованих на основі атрибутів певного об'єкту (2.44) [84]

$$p_i(l, x) = \begin{cases} 1, a_i \in P_l \\ 0, a_i \notin P_l \end{cases}, \quad (2.44)$$

де  $a_i \in A_x$  – певний атрибут об'єкту  $x$ ,  $P_l$  – множина ознак лексеми  $l$ .

Таким чином, можна задати особливості текстової розмітки типових документів, таких як закони, стандарти, положення, інструкції, наукові звіти та ін. При цьому основною вимогою є сумісність множин  $A_x$  і  $P_l$ , що дозволить ефективно використовувати операції порівняння. Для цього потрібно, щоб

первинна структура тексту  $T_{SN}$ , яка використовується при роботі методу редукції, була представлена у форматі онтології (2.1).

ЗПМ може формуватися на основі даного реєстру, результатів роботи методу онтолого-керованої рекурсивної редукції (2.45).

$$O \cup O_S \xrightarrow{G_T} T_S, \quad (2.45)$$

де  $O$  – онтологія, що містить результати методу рекурсивної редукції;

$O_S$  – онтологічний реєстр стилів;

$T_S$  – класифікаційна таксономія, що має структуру (2.46).

$$T_S = \langle X_T, R_T \rangle, \quad (2.46)$$

де  $X_T$  – множина об'єктів, що належать таксономії;

$R_T$  – множина зв'язків, що містяться в таксономії.

Множина  $X_T$  складається з об'єктів  $X_S$  початкової онтології  $O_S$ , а також множини об'єктів, що представляють класифікаційні властивості відповідних стилів (побудованих на основі атрибутів  $A'_S$  об'єктів  $X_S$ ):

$$X_T = X_S \cup X(A'_S) \quad (2.47)$$

Множина  $R_T$  містить взаємозв'язки між об'єктами  $X_T$ , що представляють стилі, і об'єктами  $X(A'_S)$ , які представляють класифікаційні характеристики об'єктів:

$$R_T = R(X_S, X(A'_S)) \quad (2.48)$$

Для таксономії  $T_S$  можна побудувати таку ЗПМ, яка буде унівалентною до неї (2.49).

$$T_S \xrightarrow{G_\Psi} \Psi_S, T_S \cong \Psi_S \quad (2.49)$$

### 2.2.3. Модель трандисциплінарної консолідації інформаційних ресурсів

Для забезпечення коректного консолідованого використання ІР створених за різними стандартами та на базі різних інтерфейсів застосовують принципи

трансдисциплінарної інтеграції політематичних інформаційних ресурсів глобального середовища [80, 92, 115].

Ідею конструктивної можливості консолідованої взаємодії мережевих ІР викладено в роботі [74]. Вона полягає в умовному розміщенні таких ресурсів в єдиному та певним чином упорядкованому цифровому просторі, спираючись на принцип трансдисциплінарності, а саме представлення усіх процесів на основі певної часткової упорядкованості множин таксономічних та операціональних властивостей онтологічних моделей ПГ. Таким чином реалізуються процедури консолідації ІР потрібних для вирішення складних задач міждисциплінарного характеру.

Така консолідована взаємодія мережевих ІР конструктивно можлива, якщо умовно розмістити їх в єдиному і упорядкованому інформаційному просторі, існування якого можливе за умови, якщо його властивості розглядати на основі категорії трансдисциплінарності.

Згідно з [86, 128, 129] інформаційно-аналітичні засоби з інтегрованими трансдисциплінарними інструментами побудови онтологій забезпечують:

- вибір та накопичення інформації, проведення її онтологічного та семантичного аналізу з подальшою побудовою тематичних словників, тезаурусів, глосаріїв;
- системно-онтологічну структурування і формально-логічне представлення інформації;
- контекстно-орієнтовану тематичну класифікацію та каталогізацію інформації;
- витяг із множини текстових документів знань, релевантних до заданої тематики;
- інтеграцію онтологічних описів, як основних компонентів методології міждисциплінарного представлення інформаційних статей;

– автоматичну побудову ієрархії термінів заданої глибини, відповідних таксономій та тезаурусів термінів, описи яких включені до файлових електронних репозиторіїв текстових документів;

– створення тематичних конспектів документів як за темами, визначеними експертами, так і за автоматично виділеними темами з можливістю автоматичного розширення заданої теми за рахунок зв'язаних тем, які можуть або визначатись автоматично під час аналізу документа, або задаватись заздалегідь у вигляді фрагмента онтології ПГ;

– візуалізацію побудованої системи понять у веб-браузерах, за рахунок чого досягається кросплатформеність представлення результатів;

– підключення через інтерфейси інтероперабельної взаємодії користувачів до інших ІС з метою розширення їх можливостей на основі використання технології обміну даними з іншими системами.

Онтологічний шаблон представлення – це вид онтології, призначений для визначення додаткових модулів натуральної системи. Ці модулі можуть бути включені в структуру системи, змінюючи її поведінку відповідно до вимог завдання.

Використання онтологічного шаблону в інтерактивному документі перетворює його з пари (2.28) у трійку (2.50).

$$\langle O, O_D, SN \rangle, \quad (2.50)$$

де  $O_D$  – онтологічний шаблон представлення.

Узагальнена інформаційна модель натуральної системи, яка підтримує онтологічні шаблони представлення, повинна мати вигляд (31)

$$P_{SN} = \sum_{i=0}^n P_{SN}^i \cup G_T(O_D), \quad (2.51)$$

де  $P_{SN}^i$  – стандартні модулі, що забезпечують основні функції обробки та відображення даних;

$G_T$  – перетворення інтерпретації онтологічного шаблону представлення.

Основним стандартним модулем такої натуральної системи є системний контролер, який контролює взаємодію між іншими модулями, оскільки вони виконують цільову функцію системи (2.30). Системний контролер забезпечує перетворення консолідації функцій окремих модулів (2.52).

$$G_C : \bigcup_{i=0}^n S_i \cup S_T(O_D) \rightarrow \check{f}, \quad (2.52)$$

де  $G_C$  – перетворення консолідації;

$S_i$  – функції стандартних модулів  $\Pi_{SN}^i$ ;

$S_T(O_D)$  – функції, отримані в результаті інтерпретації онтологічного шаблону представлення. Інтерпретація поділяється на два паралельні процеси - створення функцій обробки даних та функцій відображення даних (2.53).

$$S_T(O_D) = \sum_{x \in X_D} G_Q(x) \cup \sum_{x \in X_D} G_D(x), \quad (2.53)$$

де  $X_D$  – сукупність об'єктів, що належать до онтології  $O_D$ ,

$G_Q$  і  $G_D$  – перетворення інтерпретації, призначені для створення функцій обробки даних та представлення відповідно.

Використання перетворень у формі (2.53) означає, що натуральна система буде містити багато функцій представлення інформації, які в загальному випадку не можуть використовуватися одночасно. Отже, структура цільової функції (2.30) дозволяє користувачеві вибрати функцію відображення, яка працюватиме на даний момент. Для цього контролер системи  $\Pi_0$  повинен забезпечити додаткову функцію перемикача (2.54).

$$D_0(X, \check{x}) = \begin{cases} D_1(X), \check{x} = \check{x}^1 \\ D_2(X), \check{x} = \check{x}^2 \\ \dots \\ D_n(X), \check{x} = \check{x}^n \end{cases}, \quad (2.54)$$

де  $X$  – певний набір об'єктів (як правило, результат обробки функцій обробки інформації);

$\tilde{x}$  – команда користувача, що визначає вибір функції відображення;

$D_i$  – функції відображення, що забезпечуються системними модулями;

$\tilde{x}^i$  – маркери функцій відображення.

Використання онтологічних шаблонів представлення вимагає розширення системи генерації інтерактивних документів. Він повинен включати підсистеми, які забезпечуватимуть  $G_O$  і  $G_D$ . Для їх використання можуть знадобитися додаткові обчислювальні ресурси, особливо протягом періоду розробки, коли очікуються постійні зміни в онтологічних шаблонах представлення. Крім того, їх використання значно підвищує вимоги щодо захисту від несанкціонованого доступу до онтологій  $O_D$ . Тому використання онтологічних шаблонів представлення після закінчення періоду розробки може бути недоцільним – буде ефективніше створити інтерактивний документ із формою (2.29), щоб його натуральна система вже включала всі описані онтологічно модулі як стандартні модулі  $P_i$ .

Задача трансдисциплінарної консолідації інформації вирішується шляхом перетворення розширення контекстами об'єктів онтології, отриманої в результаті роботи методу рекурсивної редукції (2.55).

$$W^T \rightarrow X_t(W^T), \quad (2.55)$$

де  $W^T$  – контекст;

$X_t$  – концепт, що міститься в контексті.

В подальшому такі онтології використовуються як вхідні дані для функції пошуку (2.56).

$$Q_s(H, x) = \left\{ H \left\langle \left\{ V(l) \times V(x) \right\} \right\rangle \right\}, \quad (2.56)$$

де  $H$  – індекс, отриманий в результаті індексації масиву наративів мережевих документів за допомогою спеціалізованої функції індексації  $Q_H$ ;

$V(l), V(x)$  – ідентифікатори лексеми  $l$  і об'єкта  $x$ , що належать онтології  $O$ .



Функція пошуку дозволяє формувати зв'язки між контекстами всіх лексичних одиниць безлічі мережевих нарративних документів. Дана операція є основною для побудови функції контекстної зв'язності (2.57)

$$Q_C(L_x, P) = \bigcup_{O \in P} (Q_S(Q_H(O), L_x)), \quad (2.57)$$

де  $P$  – множина онтологій  $O$ , що описують одну або кілька ПГ;

$L_x$  – текстове представлення контексту лексичної одиниці  $x$ , що належить таксономії  $T$ ;

Основним недоліком функції (2.57) є формування в результаті її застосування, великої неупорядкованої множини документів, обробка якої може викликати труднощі. Для рішення цієї проблеми пропонується два способи [90]:

- 1) Задання на множині результатів відношення релевантності  $R_{rel}$ :

$$T_1 R_{rel} T_2 \Rightarrow card(L_{T_1} \cap L_x) > card(L_{T_2} \cap L_x) \quad (2.58)$$

де  $L_{T_1}, L_{T_2}$  – текстові представлення документів  $T_1, T_2$ ;

- 2) Виключення з результату документів з недостатньою релевантністю, що досягається шляхом застосування в функції (2.57) операції перетину замість операції об'єднання.

Для здійснення такої операції необхідно виконати перетворення з допомогою функції:

$$Q_{II}(P) = \bigcup_{x \in X_P} \{Q_P(x)\}, \quad (2.59)$$

де  $P$  – множина онтологій;

$X_P$  – множина об'єктів, що належать об'єднанню онтологій  $\bigcup_{O \in P} O$ .

Застосування функції контекстної зв'язності утворює множину об'єктів з різних онтологій, представляючи таким чином гіпервідношення між певними об'єктами. З допомогою даних гіпервідношень можна побудувати трансдисциплінарне представлення  $O'$  множини онтологій  $P$ :

$$P \xrightarrow{Q_{\pi}} \langle \bigcup_i X_i, \bigcup_i R_i \cup Q_{\pi}(P), \bigcup_i F_i \rangle \quad (2.60)$$

Перетворення (2.60) задає найбільш повне представлення наявної в множині онтологій інформації. Однак, найчастіше виникає необхідність виконати представлення однієї вибраної онтології  $O$ . В такому випадку (2.59) необхідно змінити:

$$Q_{TO}(O, P) = \bigcup_{x \in X_O} \{Q_P(x)\}, \quad (2.61)$$

де  $P$  – множина онтологій;

$X_O$  – множина об'єктів, що належать онтології  $O$ .

Множина індексів  $H$ , що формується на основі застосування до множини  $P$  функції індексації  $Q_H$ , формує індексну зону  $\tilde{H}$  всіх наративів мережевих документів  $O_{nd}$ :

$$\tilde{H} = \{\{H\} \times T\} \quad (2.62)$$

Таким чином, функцію контекстної зв'язності можна представити у вигляді:

$$Q_P(l) = \tilde{H} \times P \times \tilde{T} \times \mathfrak{R}^3 \quad (2.63)$$

Функція контекстної зв'язності створює умови для формування наративного дискурсу на основі семантико-лексичної і концептографічної обробки всіх наративів мережевих документів  $O_{nd}$ .

Таким чином умови щодо формування наративного дискурсу на основі семантико-лексичної і концептографічної обробки всіх наративів мережевих документів  $O_{nd}$  визначаються функцією контекстної зв'язності.

Тобто трансдисциплінарна структуризація мережевих наративів та документів, визначає функціональні умови щодо виявлення контекстної зв'язності між усіма концептами-поняттями, що створюють термінополе експерта і, як наслідок, визначають їх взаємозв'язки між собою. Тоді усі вузли онтографу, що відображають цю зв'язність, утворюють множину усіх станів інтерактивної

взаємодії з контекстами множини мережевих документів, з якими взаємодіє експерт у процесі розв’язання конкретної задачі.

## 2.3. Формалізація документів різних стилів на основі рекурсивної редукції

### 2.3.1. Класифікація функціональних стилів

Мовознавці одностайно констатують наявність у мові різних функціональних стилів. Їх природа зумовлена об’єктивною реальністю. Проте, принципи виокремлення та класифікації функціональних стилів значною мірою залежать від погляду того чи іншого дослідника. Сьогодні існують різні підходи до класифікації функціональних стилів. Їх кількість, назви і склад підстилів не збігаються.

Найпоширенішою у науковій літературі є, на сьогодні, класифікація, згідно з якою виділяють п’ять основних стилів мови: *науковий, офіційно-діловий, публіцистичний, художній та розмовний стилі* (М. Пилинський, О. Пономарів, М. Пентилюк, Н. Ботвина, С. Єрмоленко, Н. Озерова) [130]. Наведена класифікація опирається на функціонально-стилістичну концепцію академіка В. Виноградова, який пов’язує диференціацію стилів із суспільними функціями мови і може стосуватися практично всіх високорозвинених мов.

У сучасній функціональній стилістиці існують класифікації, які визначають таку систему стилів: *художній, науковий, офіційно-діловий, розмовний, публіцистичний та інформаційний стилі*. Останній виходить за межі власне публіцистичного стилю (Д. Баранник, Ю. Арешенков). «В інформаційному стилі реалізується мовна функція повідомлення, інформування, тоді як у власне публіцистичних текстах актуалізується функція впливу, або апелятивна [131]». Цей різновид спеціального мовлення діє у сфері масової комунікації (преса, радіо, телебачення, інформаційні агентства, редакції, творчі об’єднання) та має на меті повідомити про найважливіші явища навколишньої дійсності.

Деякі дослідники розглядають як самостійний *епістолярний стиль*. В праці “Стиль і час. Хрестоматія” [133] наведено такий поділ стилів в українській

літературній мові: *стиль художньої літератури, публіцистичний стиль, науковий, офіційно-діловий та епістолярний*.

Останнім часом мовознавці Н. Бабич, В. Задорожний, Н. Дзюбишина-Мельник додають до цих класифікацій ще одну, яка містить у своєму складі *конфесійний стиль* [135], що пов'язаний з особливою сферою людського буття — релігією. Зараз можна говорити про сформованість цього стилю як окремого різновиду української мови і вважати його повноцінним складником її функціонально-стильової сфери.

Сьогодні багато уваги приділяють дослідженню теорії та історії *ораторського стилю*. Суттєвий внесок у це питання зробили праці Мацько Л. І. [137].

Кожний із функціональних стилів має свої підстили, або жанрові різновиди. У цьому плані думки мовознавців теж не є однотайними.

У межах *наукового стилю (НС)* дослідниками виділяються, насамперед 1) *власне науковий (академічний) підстиль* із жанрами текстів (монографія, рецензія, стаття, наукова доповідь, повідомлення, курсова й дипломна роботи, реферат, тези); 2) *науково-довідковий* (довідники, енциклопедії, каталоги); 3) *науково-навчальний (дидактичний) підстиль* із жанрами текстів (підручники, посібники, лекції, бесіди); 4) *науково-популярний підстиль* (науково-популярна література). Деякі мовознавці вважають, що україномовний власне науковий підстиль майже не реалізується. За їх переконаннями, функції власне наукового підстилю виконує в україномовному спілкуванні *науково-технічний мікростиль* (А. Коваль, Н. Непийвода) [139].

Дослідниця *офіційно-ділового стилю (ОДС)* А. Марахова виділяє такі його підстили: 1) *законодавчий* (закони, укази, статути, постанови); 2) *дипломатичний* (декларації, міжнародні договори, конвенції, комюніке, меморандуми, ноти, угоди, ультиматуми, міжнародні заяви, акти); 3) *адміністративно-канцелярський* (статути, інструкції, постанови, накази, розпорядження, договори, контракти, заяви, автобіографії, звіти, службові листи, доповідні, характеристики, свідоцтва, розписки та ін.). Деякі мовознавці сюди відносять *юридичний підстиль*, який використовується в юриспруденції (судочинство, дізнання, розслідування, арбітраж) (М. Зубков).

*Публіцистичний стиль (ПС)*, на думку Г. Колесник, характеризується такими підстилями: 1) *власне публіцистичним* (повідомлення, хроніка, прокламації, відкриті листи, передові статті, огляди міжнародного життя); 2) *науково-публіцистичним* (великі праці на теми сучасної політики, економіки та суспільних відносин, літературно-критичні статті, наукові та літературно-мистецькі огляди, рецензії); 3) *художньо-публіцистичним* (нариси, фейлетони, памфлети, мемуари, політичні доповіді). Як окремі підстилі ПС С. Єрмоленко розглядає *мову преси* (інтерв'ю, репортаж, хроніка, нарис); *мову радіо і телебачення* (мова дикторів, ведучих радіо і телепередач); *усну публічну мову* (промови, виступи на зборах, мітингах). Д. Баранник виділяє *церемоніальний різновид ПС*.

За родами і жанрами літератури поділяється на підстилі *художній стиль (ХС)*: 1) *епічний, прозовий* (епопея, казка, роман, повість, оповідання, байка, новела, художні мемуари, нарис); 2) *ліричний, поетичний* (поема, балада, пісня, гімн, елегія, епіграма); 3) *драматичний* (драма, трагедія, комедія, мелодрама, водевіль); 4) *комбінований* (ліро-епічний твір, ода, художня публіцистика, драма-фесерія, усмішка) [141].

У *розмовному стилі (РС)* виділяють наступні підстилі: 1) *розмовно-побутовий* і 2) *розмовно-офіційний*.

Дослідниця *інформаційного стилю (ІС)* Н. Леонова виокремлює 1) *власне інформаційний* та 2) *аналітичний (коментуючий, об'єктивно-інформаційний)*. Дискусійним залишається існування *рекламного підстилю*.

*Конфесійний стиль (КС)* на думку дослідника І. Грималовського, варто поділити на такі підстилі: 1) *сакральний* (Святе письмо, канонічні молитви); 2) *літургійний* (літургійні тексти); 3) *проповідницький* (релігійні проповіді); 4) *науково-конфесійний* (наукові праці з богослов'я); 5) *художньо-конфесійний* (авторські молитви, переспіви біблійних текстів, наприклад, Шевченкові “Давидові псалми”).

Дослідники *епістолярного стилю (ЕС)* К. Ленець, М. Пилинський, В. Німчук, Б. Шарпило заявляють про наявність у цьому стилі двох підстилів: 1) *офіційного, службового* і 2) *неофіційного, приватного*. “Залежно від теми листа, його

призначення і взаємин кореспондентів, — пише К. Ленець, — листи можна поділити на родинно-побутові; інтимно-товариські; приватно-ділові; офіційно-ділові” [133].

Г. Апресян розрізняє такі види ораторського стилю (ОС): 1) *громадсько-політичне ораторське мистецтво* (лекції на громадсько-політичні теми, політичні промови, огляди, виступи на мітингах, агітаційні промови); 2) *академічне ораторське мистецтво* (наукова доповідь); 3) *судове ораторське мистецтво*; 4) *промови з нагоди громадських урочистостей* (ювілейна промова або поздоровлення, тост); 5) *церковні проповіді*.

Отже, в сучасних умовах функціонально-стильова сфера мови постійно уточнюється: переосмислюються традиційні стилі, відновлюються ті, яких за певних умов не визнавали, формуються нові.

Узагальнена структура функціональних стилів має наступний вигляд:

## **1. ОСНОВНІ СТИЛІ**

### 1.1 Науковий

*1.1.1 Власне науковий (академічний);*

*1.1.2 Науково-довідковий;*

*1.1.3 Науково-навчальний (дидактичний);*

*1.1.4 Науково-популярний;*

### 1.2 Офіційно-діловий

*1.2.1 Законодавчий;*

*1.2.2 Дипломатичний;*

*1.2.3 Адміністративно-канцелярський;*

*1.2.4 Юридичний*

### 1.3 Публіцистичний

*1.3.1 Власне публіцистичний;*

*1.3.2 Науково-публіцистичний;*

*1.3.3 Художньо-публіцистичний;*

*1.3.4 Мова преси;*

*1.3.5 Мова радіо і телебачення;*

*1.3.6 Усна публічна мова;*

### *1.3.7 Церемоніальна мова*

## 1.4 Художній

### *1.4.1 Епічний, прозовий;*

### *1.4.2 Ліричний, поетичний;*

### *1.4.3 Драматичний;*

### *1.4.4 Комбінований*

## 1.5 Розмовний

### *1.5.1 Розмовно-побутовий;*

### *1.5.2 Розмовно-офіційний*

## **2. ІНШІ СТИЛІ**

### 2.1 Інформаційний

#### *2.1.1 Власне інформаційний;*

#### *2.1.2 Аналітичний*

### 2.2 Конфесійний

#### *2.2.1 Сакральний;*

#### *2.2.2 Літургійний;*

#### *2.2.3 Проповідницький;*

#### *2.2.4 Науково-конфесійний;*

#### *2.2.5 Художньо-конфесійний*

### 2.3 Епістолярний

#### *2.3.1 Офіційний, службовий;*

#### *2.3.2 Неофіційний, приватний*

### 2.4 Ораторський

#### *2.4.1 Громадсько-політичне ораторське мистецтво;*

#### *2.4.2 Академічне ораторське мистецтво;*

#### *2.4.3 Судове ораторське мистецтво;*

#### *2.4.4 Урочисті промови;*

#### *2.4.5 Церковні проповіді.*

### **2.3.2. Шкала ступеню формалізованості документу**

Для аналізу функціональних стилів і підстилів з точки зору ефективності їх автоматизованої обробки, необхідно визначити ознаки, наявність яких у текстах

конкретного стилю (підстилю) підвищує ефективність обробки, і, на основі цього сформувані критерії аналізу.

Обробка здійснюється за допомогою інструментального засобу «Рекурсивний редуктор». Роботу рекурсивного редуктора, як інструменту структуризації текстів, побудовано на принципі автоматичного виділення з вхідного тексту, що обробляється:

- 1) Сутностей. Сутності мають бути явно вказані в тексті. Наприклад: ім'я людини, назва тварини, міста, вулиці, назва товару, місяць, рік, і т. ін.;
- 2) Відношень "вище-нижче", "клас-підклас", "клас-об'єкт ". Наприклад: присутня категорія "академіки" і, власне, академіки, що до неї належать;
- 3) Атрибутів об'єктів, наприклад: автомобіль, колір: зелений;
- 4) Контекстів, тобто фрагментів вхідного тексту, що стосуються того чи іншого об'єкту.

Допоміжні ознаки, які можуть зустрічатись у вхідному тексті, і застосовуватись в процесі його обробки:

- 1) Загальна структура документу (зміст, заголовки);
- 2) Роздільники. В тексті мають бути чіткі межі різних смислових фрагментів. Наприклад, абзаци, що стосуються конкретних людей;
- 3) Списки. Редуктор виділяє однорівневі описи сутностей, що розміщені послідовно. Наприклад: список людей;
- 4) Спеціалізовані об'єкти: таблиці, рисунки. Такі об'єкти можуть бути джерелом важливої інформації, але в залежності від їх структури (складні заголовки таблиць, багатострочні підписи до рисунків з вказівкою різних частин, типу (рис 1, а; рис 1, б)) можуть вимагати складних процедур аналізу.

Способи обробки документів редуктором:

- 1) Аналіз структури (зміст, списки та ін.) документу;
- 2) Регулярні конструкції;
- 3) Застосування словників. Якщо вже існує тезаурус сутностей (наприклад, список імен і прізвищ) то обробку можна виконувати достатньо просто; якщо в



тексті є стандартизовані назви, але не існує тезаурусу для них – то гірше; якщо ж в тексті широко застосовується синонімія, то використання словників, як правило, є неефективним.

Отже, критеріями аналізу стилів (підстилів) на ефективність обробки редуктором будуть наявність у них і якість оформлення: сутностей; відношень; атрибутів; контекстів; змісту (рубрикації); роздільників; списків; спеціалізованих об'єктів.

Складність оцінки визначених критеріїв полягає в тому, що необхідно оцінити не лише наявність в тексті даного підстилю кожного критерію, але і якість його оформлення, що має високу важливість для автоматичної обробки. Тому запропоновано оцінювати критерії по такій лінгвістичній шкалі:

- низька наявність – якщо в тексті художнього стилю майже ніколи не присутні списки, немає необхідності оцінювати їх якість в такому тексті. В цьому випадку буде присвоєно значення «низька наявність»;

- низька важливість – наприклад, при оформленні поетичних творів широко застосовуються роздільники (між строфами), але тут вони грають роль інструменту, що більше слугує зручності читання, аніж логічному поділу тексту на частини, і тому роздільники в даному випадку мають низьку важливість. Або ілюстрації в художньому тексті, на відміну від наукового;

- висока складність обробки – наприклад, в документах офіційно-ділового стилю часто застосовують списки, але в межах одного пункту списку зустрічається багато роздільників, що викликає складність при автоматичній обробці;

- допустима складність – бувають випадки, коли за певних умов редуктором можна обробляти складні тексти, тоді складність вважається допустимою або прийнятною. Наприклад, наявність спеціальних словників може спростити обробку тексту;

- висока якість – параметр, що демонструє найбільш сприятливі умови для автоматичної обробки тексту.

В числовому вираженні шкала матиме наступний вигляд:

- низька наявність, = 1
- низька важливість, = 1
- висока складність, = 4
- допустима складність, = 6
- висока якість, = 10

Результат експертної оцінки підстилів відображено в Додатку Г.

Ранжування стилів здійснено за допомогою СППР Альтернатива. Вибір системи переваг проілюстровано на рис. 2.4.

Враховуються властивості Рівень ранжування 1 (звичайний)

Вибір	Ім'я	Ваг. Коеф.	Опт (max/min)	Способи задання вагових коефіцієнтів		
				Бальна шкала (10)	Лінгвістична шкала	Ранжування
<input checked="" type="checkbox"/>	Сутності (+)	0.179	маг; ▼	5 ▼	Абсолютна важливість ▼	1 ▼
<input checked="" type="checkbox"/>	Відношення (+)	0.143	маг; ▼	5 ▼	Значна важливість ▼	1 ▼
<input checked="" type="checkbox"/>	Атрибути (+)	0.143	маг; ▼	5 ▼	Значна важливість ▼	1 ▼
<input checked="" type="checkbox"/>	Контексти (+)	0.143	маг; ▼	5 ▼	Значна важливість ▼	1 ▼
<input checked="" type="checkbox"/>	Зміст, рубрикація (+)	0.107	маг; ▼	5 ▼	Середня важливість ▼	1 ▼
<input checked="" type="checkbox"/>	Роздільники (+)	0.107	маг; ▼	5 ▼	Середня важливість ▼	1 ▼
<input checked="" type="checkbox"/>	Списки (+)	0.107	маг; ▼	5 ▼	Середня важливість ▼	1 ▼
<input checked="" type="checkbox"/>	Спеціалізовані об'єкти (+)	0.071	маг; ▼	5 ▼	Слабка важливість ▼	1 ▼
8/8	Система переваг		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Рисунок 2.4 – Система переваг критеріїв

Результат проведеного ранжування, представлений в таблиці 2.2, є шкалою ефективності обробки рекурсивним редуктором документів різних стилевих груп. По суті вона є шкалою ступеня формалізованості документів різних стилів.

Таблиця 2.1 – Результат ранжування

#	Елементи	Значення	Критерії								
			Сутності	Відношення	Атрибути	Контексти	Зміст, рубрикація	Роздільники	Списки	Спеціалізовані об'єкти	
1	Науково- довідковий	1	висока якість	висока якість	допустима складність	висока якість	висока якість	висока якість	висока якість	висока якість	допустима складність
2	Законодавчий	0.769	допустима складність	допустима складність	допустима складність	висока якість	висока якість	висока якість	допустима складність	допустима складність	низька наявність
3	Науково- навчальний	0.724	допустима складність	допустима складність	допустима складність	допустима складність	висока якість	висока складність	допустима складність	допустима складність	допустима складність
4	Власне науковий	0.575	висока складність	висока складність	висока складність	висока складність	допустима складність	висока складність	допустима складність	допустима складність	допустима складність
5	Адміністративно- канцелярський	0.572	висока складність	допустима складність	висока складність	допустима складність	висока складність	висока складність	допустима складність	допустима складність	низька наявність
6	Власне інформаційний	0.507	допустима складність	висока складність	висока складність	висока складність	висока складність	висока складність	низька наявність	низька наявність	низька наявність
7	Дипломатичний	0.475	допустима складність	низька наявність	висока складність	висока складність	допустима складність	низька важливість	висока складність	низька наявність	низька наявність
8	Науково- популярний	0.411	висока складність	низька наявність	висока складність	висока складність	допустима складність	низька важливість	низька наявність	низька наявність	низька важливість
9	Юридичний	0.41	допустима складність	висока складність	низька наявність	висока складність	висока складність	низька важливість	низька наявність	низька наявність	низька наявність

Продовження таблиці 2.1

10	Науково-публіцистичний	0.396	висока складність	низька наявність	висока складність	висока складність	висока складність	низька важливість	низька наявність	низька наявність
11	Церемоніальна мова	0.393	висока складність	низька наявність	допустима складність	висока складність	низька наявність	низька наявність	низька наявність	низька наявність
12	Науково-конфесійний	0.38	допустима складність	низька наявність	низька наявність	висока складність	допустима складність	низька важливість	низька наявність	низька наявність
13	Неофіційний	0.371	висока складність	низька наявність	низька наявність	висока складність	низька наявність	низька важливість	висока якість	низька наявність
14	Академічне ораторське мистецтво	0.366	висока складність	низька наявність	висока складність	висока складність	низька наявність	низька наявність	низька наявність	низька наявність
15	Офіційний	0.349	висока складність	низька наявність	низька наявність	висока складність	низька наявність	низька важливість	допустима складність	низька наявність
16	Власне публіцистичний; мова преси	0.334	висока складність	низька наявність	низька наявність	висока складність	висока складність	низька важливість	низька наявність	низька наявність
17	Епічний, прозовий; ліричний, поетичний	0.305	висока складність	низька наявність	низька важливість	висока складність	низька важливість	низька важливість	низька наявність	низька наявність
18	Розмовно-побутовий	0.305	висока складність	низька наявність	низька важливість	висока складність	низька наявність	низька наявність	низька наявність	низька наявність

Продовження таблиці 2.1

19	Драматичний	0.305	висока складність	низька наявність	низька наявність	висока складність	низька важливість	низька важливість	низька важливість	низька наявність
20	Сакральний; літургійний; комбінований	0.305	висока складність	низька наявність	низька наявність	висока складність	низька важливість	низька важливість	низька наявність	низька наявність
21	Художньо- конфесійний; мова радіо і телебачення; художньо- публіцистичний; аналітичний	0.305	висока складність	низька наявність	низька наявність	висока складність	низька наявність	низька важливість	низька наявність	низька наявність
22	Церковні проповіді; урочисті промови; громадсько- політичне ораторське мистецтво; судове ораторське мистецтво; проповідницький; розмовно-офіційний	0.305	висока складність	низька наявність	низька наявність	висока складність	низька наявність	низька важливість	низька наявність	низька наявність

### 2.3.3. Оцінка еквівалентності документів різних стилів з точки зору автоматизованої обробки

Обробивши таблицю оцінки підстилів за допомогою програми Конфор, можна точно виявити підстилі, які є абсолютно еквівалентними за обраними критеріями. Результат такої обробки представлено на рис. 2.5.

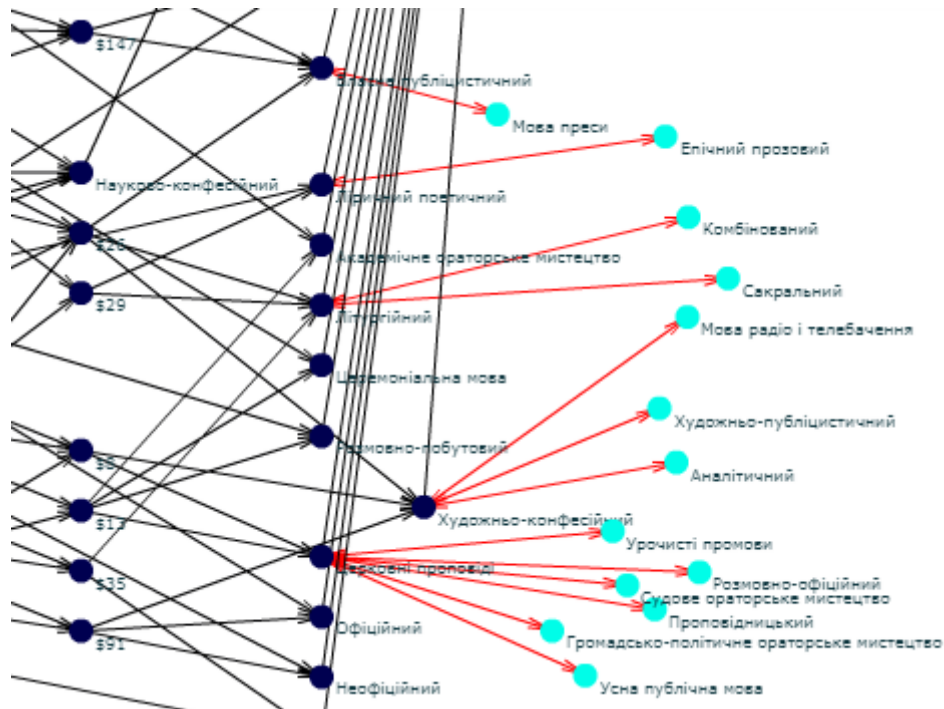


Рисунок 2.5 – Фрагмент онтології еквівалентних підстилів

Еквівалентними підстилями виявились:

- 1) Власне публіцистичний та мова преси;
- 2) Епічний прозовий та ліричний поетичний;
- 3) Сакральний літургійний та комбінований;
- 4) Художньо-конфесійний, мова радіо і телебачення, художньо-публіцистичний та аналітичний;
- 5) Церковні проповіді, урочисті промови, громадсько-політичне ораторське мистецтво, судове ораторське мистецтво, проповідницький та розмовно-офіційний.

Таким чином, визначено групи підстилів, які можуть виступати як один підстиль, оскільки з точки зору автоматичної обробки різниці між ними немає.

## 2.4. Висновки за розділом 2

Розроблено концептуальну модель трансдисциплінарного представлення мережевих інформаційних ресурсів.

Обґрунтовано створення програмно-інформаційного засобу трансдисциплінарного представлення інформаційних ресурсів різних стилів.

Представлено модель онтолого-керованої ідентифікації інформації, в результаті чого отримав розвиток метод рекурсивної редукції. Показано механізм формування онтологій документів на основі репрезентації їх таксономій, зв'язків між їх контекстами та множинного їх представлення, що забезпечить ефективний пошук інформації у великих масивах різнорідних джерел.

Проведено класифікацію функціональних стилів мови. Розроблено шкалу ступеню формалізованості документу. Проведено оцінку еквівалентності документів різних стилів з точки зору автоматизованої обробки.

Застосування онтологічного класифікатора стилів дозволяє підвищити ефективність обробки масивів різнорідних документів, а включення до класифікатора описів представлення текстів посилань на інші документи дозволяє створювати на їх основі системи взаємозв'язаних онтологій, що представляють певну ПГ.

Результати досліджень, наведені в другому розділі, опубліковано в роботах [10, 83, 84, 90, 128, 132, 134, 136, 138, 140, 142, 143].

### РОЗДІЛ 3. АРХІТЕКТУРА ПРОГРАМНОЇ СИСТЕМИ ТРАНСДИСЦИПЛІНАРНОГО ПРЕДСТАВЛЕННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ РІЗНИХ СТИЛІВ

З точки зору програмної інженерії будь-яка програмна система являє собою зв'язні набори описів, що представимо у форматі математичних моделей, формалізованих висловлювань та виразів і технік моделювання [109, 144].

Для опису архітектури програмної системи трансдисциплінарного представлення ІР різних стилів обрано об'єктно-орієнтовану парадигму проектування програмних систем, що базується на множині класів і зразках взаємодії, що визначають, кооперації цих класів у реалізації певних сценаріїв [145]. Дана парадигма реалізується за допомогою уніфікованої мови моделювання UML, яка забезпечує візуалізацію, специфікацію, конструювання й документування (ВСКД) об'єктно-орієнтованих систем.

Програмна система трансдисциплінарного представлення ІР різних стилів призначена для автоматизованої обробки різностильових документів глобального середовища та подання їх у формі, придатній для подальшого перетворення в активний формат інтерактивних баз знань. Практична реалізація такої технології здійснюється на основі WEB-орієнтованого сервісу Когнітивна інформаційна технологія (КІТ) «Полієдр» [146].

Структура ПС такого роду включає в себе наступні моделі [109, 144]:

- 1) інформаційна модель;
  - a) інформаційна модель підсистеми класифікації ПМТ за стилями;
  - b) інформаційна модель підмножини модулів КІТ;
- 2) функціонально-компонентна модель;
  - a) модель поведінки системи;
    - i) модель варіантів використання системи;
    - ii) модель активності системи;
    - iii) модель взаємодії системи;
  - b) модель структури системи;



- c) модель структури програмних сутностей;
- d) схема компонентів програмної системи;

архітектура системи трансдисциплінарного представлення IP різних стилів.

### 3.1. Інформаційна модель програмної системи трансдисциплінарного представлення різностильових IP

Інформаційну модель програмної системи трансдисциплінарного представлення IP різних стилів  $PC_{pc}$  можна виразити як [109, 134]:

$$PC_{pc} = \sum_{i=1}^n M_{PC_i} \cup IT_{Pol}, \quad (3.1)$$

де  $M_{PC_i}$  – скінченна сукупність програмних модулів програмної системи;

$IT_{Pol}$  – когнітивна інформаційна технологія (КІТ) «Поліедр».

Когнітивна інформаційна технологія (КІТ) «Поліедр», в свою чергу, також має значну кількість багатofункціональних модулів різного призначення. В роботі системи трансдисциплінарного представлення IP різних стилів використовується тільки певна сукупність модулів, що релевантні задачі перетворення ПМТ різних стилів у формат інтерактивних документів. Таким чином формула (3.1) перетворюється в (3.2)

$$PC_{pc} = \sum_{i=1}^n M_{PC_i} \cup \sum_{i=1}^m IT_{Pol_i} \quad (3.2)$$

Сукупність модулів системи трансдисциплінарного представлення IP різних стилів, що не входять до складу базової структури «Поліедр», можна розглядати як окрему підсистему – підсистему класифікації ПМТ за стилями  $M_{KC} = \sum_{i=1}^n M_{PC_i}$ .

Дана сукупність матиме вигляд (3.3)

$$M_{KC} = \{M_{IC}, M_{\PhiЗПМ}, M_{КТ}, M_{ІШ}\}, \quad (3.3)$$

де  $M_{IC}$  – модуль інтерпретації онтології стилів;

$M_{\PhiЗПМ}$  – модуль формування ЗПМ;

$M_{КТ}$  – модуль класифікації вхідного тексту за стилем;

$M_{ІШ}$  – модуль інтерпретації шаблонів представлення.

Кожен з модулів виконує наступні функції:

**Модуль інтерпретації онтології стилів**

$$S_{IC} = \langle S_1^{IC}, S_2^{IC}, S_3^{IC}, S_4^{IC}, S_5^{IC} \rangle, \quad (3.4)$$

де  $S_1^{IC}$  – здійснення спеціалізованої процедури створення кінцевих множин функцій інтерпретації  $F$  і додаткових обмежень  $R_s$ ;

$S_2^{IC}$  – класифікація елементів  $F$  і  $R_s$  згідно з їх ролями і підготовка для формування правил;

$S_3^{IC}$  – формування функцій інтерпретації;

$S_4^{IC}$  – формування функцій перетворення;

$S_5^{IC}$  – формування правил і представлення їх в форматі рекурсивного редуктора.

**Модуль формування ЗПМ**

$$S_{\PhiЗПМ} = \langle S_1^{\PhiЗПМ}, S_2^{\PhiЗПМ}, S_3^{\PhiЗПМ} \rangle, \quad (3.5)$$

де  $S_1^{\PhiЗПМ}$  – формування множини рецепторів пірамідальної мережі, тобто вершин, що не мають вхідних зв'язків;

$S_2^{\PhiЗПМ}$  – реконфігурація структури зростаючої пірамідальної мережі шляхом об'єднання рецепторів та формування з них концепторів із використанням певних правил;

$S_3^{\PhiЗПМ}$  – виділення понять у структурі зростаючої пірамідальної мережі шляхом розмітки вершин мережі з метою отримання висновків у нотації логіки висловлювань для елементів мережі.

**Модуль класифікації вхідного тексту за стилем**

$$S_{KT} = \langle S_1^{KT}, S_2^{KT}, S_3^{KT}, S_4^{KT}, S_5^{KT} \rangle, \quad (3.6)$$

де  $S_1^{KT}$  – зчитування класифікаційних ознак з онтології стилів;

$S_2^{KT}$  – порівняння отриманих ознак з ознаками наявними у вхідних ПМТ;

$S_3^{KT}$  – побудова класифікаційної таксономії;

$S_4^{KT}$  – запуск модулю формування ЗПМ;

$S_5^{KT}$  – інтерпретація результатів роботи модулю ЗПМ.

**Модуль інтерпретації шаблонів представлення**

$$S_{III} = \langle S_1^{III}, S_2^{III}, S_3^{III} \rangle, \quad (3.7)$$

де  $S_1^{III}$  – зчитування файлів шаблонів;

$S_2^{III}$  – формування внутрішнього представлення для предикатів, з яких формуються шаблони;

$S_3^{III}$  – формування внутрішнього представлення для правил ідентифікації інформації;

В процесі обробки різностильових документів і подання їх у формі інтерактивних документів задіяні також вже існуючі модулі КІТ «Поліедр», а саме:

$$IT_{Pol_i} = \{M_{BO}, M_{PO}, M_{PP}, M_{ТДП}\}, \quad (3.8)$$

Де  $M_{BO}$  – модуль відображення онтологій;

$M_{PO}$  – модуль редагування онтологій;

$M_{PP}$  – модуль рекурсивної редукції;

$M_{ТДП}$  – модуль трансдисциплінарного представлення.

Функції, що виконуються представленими модулями:

**Модуль відображення онтологій**

$$S_{BO} = \langle S_1^{BO}, S_2^{BO}, S_3^{BO}, S_4^{BO} \rangle, \quad (3.9)$$

де  $S_1^{BO}$  – об’єктне відображення онтологій;

$S_2^{BO}$  – відображення онтологій у вигляді онтографу;

$S_3^{BO}$  – представлення списку об’єктів онтології у вигляді таблиці;

$S_4^{BO}$  – відображення онтологічної призми.

**Модуль редагування онтологій**

$$S_{PO} = \langle S_1^{PO}, S_2^{PO}, S_3^{PO}, S_4^{PO}, S_5^{PO}, S_6^{PO} \rangle, \quad (3.10)$$

де  $S_1^{PO}$  – зчитування і запис онтологій в форматі XML;

$S_2^{PO}$  – додавання і видалення об’єктів;

$S_3^{PO}$  – додавання і видалення зв’язків між концептами;

$S_4^{PO}$  – редагування атрибутів об’єктів;

$S_5^{PO}$  – автоматичне розміщення об’єктів в робочій області;

$S_6^{PO}$  – фільтрація множини об’єктів, що відображаються.

**Модуль рекурсивної редукції**

$$S_{PP} = \langle S_1^{PP}, S_2^{PP}, S_3^{PP}, S_4^{PP}, S_5^{PP} \rangle, \quad (3.11)$$

де  $S_1^{PP}$  – виконання попереднього аналізу вхідного документу;

$S_2^{PP}$  – виконання лексичного аналізу;

$S_3^{PP}$  – формування ієрархії об’єктів згідно з наданою текстом інформацією;

$S_4^{PP}$  – формування атрибутів об’єктів;

$S_5^{PP}$  – формування онтології.

### **Модуль трансдисциплінарного представлення**

$$S_{TДП} = \langle S_1^{TДП}, S_2^{TДП}, S_3^{TДП}, S_4^{TДП}, S_5^{TДП} \rangle, \quad (3.12)$$

де  $S_1^{TДП}$  – контекстна зв'язка концептів різних онтологій;

$S_2^{TДП}$  – застосування функції пошуку;

$S_3^{TДП}$  – оптимізація множини релевантних документів;

$S_4^{TДП}$  – індексація інформаційних ресурсів;

$S_5^{TДП}$  – представлення інформації з допомогою онтологічної призми;

## **3.2. Модель поведінки системи**

Модель поведінки системи наводиться для представлення взаємодій, відносин та залежностей компонентів, з яких складається система трансдисциплінарного представлення ІР різних стилів.

Модель поведінки системи може бути представлена структурою:

$$M_D = \langle d_{use}, d_{act}, d_{seq} \rangle, \quad (3.13)$$

де  $d_{use}$  – UML-діаграма варіантів використання системи;

$d_{act}$  – UML-діаграма активності системи;

$d_{seq}$  – UML-діаграма взаємодії системи.

### **3.2.1. Варіанти використання системи**

UML-діаграму варіантів використання [147, 148] програмної системи трансдисциплінарного представлення ІР різних стилів представлено на рис. 3.1.

Дана діаграма відображає учасників процесу структурування різностильових документів і їх основні можливості. Так, в рамках системи можуть діяти «Користувач», «Адміністратор» і «Розробник».



Рисунок 3.1 – Варіанти використання системи

Для отримання доступу до системи трансдисциплінарного представлення ІР різних стилів користувач має бути авторизованим. Це дає йому право на використання закритих додатків (таких як, Confor та Індексатор) і локальних сервісів (таких як Рекурсивний редуктор). Таким чином, авторизований користувач може:

- 1) здійснювати класифікацію тексту за стилем, використовуючи функцію додатку Confor;
- 2) застосовувати рекурсивну редукцію для отримання онтологій. Для коректного виконання редукції необхідно мати базу правил, що створюється на основі обробки онтологічного реєстру стилів адміністратором системи;
- 3) переглядати інформацію в одному зі зручних режимів (об'єктне відображення, режим онтографу, табличний режим, режим онтологічної призми);
- 4) формувати трансдисциплінарну інтеграцію ІР. В середовищі КІТ «Поліедр» така інтеграція здійснюється за допомогою окремого додатку «Пошукова призма».

Адміністратор системи:

- 1) формує і наповнює онтологічний реєстр стилів. Таксономічна структура онтологічного реєстру включає 9 класів (власне стилів), 35 підкласів (підстилів),

довільну кількість об'єктів в кожному підкласі, а також атрибути кожного об'єкту, що в свою чергу можуть бути представлені у вигляді складних ієрархічних структур або таблиць. Тому адміністратор наповнює реєстр поступово, кожного разу додаючи об'єкти та їх атрибути, необхідні для вирішення конкретної задачі;

2) формує онтологічні шаблони обробки,

3) здійснює індексацію ІР. В окремих випадках можливість керування індексацією може надаватися також користувачу. Процес індексації передбачає вибір ІР тематично релевантних поставленій задачі та внесення їх у спеціальне вікно керування індексаторами.

Розробник, який має доступ до всіх функцій системи, створює онтологічні шаблони представлення. Оскільки розробник створює, модифікує і контролює роботу всіх модулів системи, він може ситуативно включати в роботу системи додаткові модулі, для забезпечення необхідної поведінки системи.

### **3.2.2. Активність системи**

Алгоритм роботи з системою трансдисциплінарного представлення ІР різних стилів показує діаграма активності [147, 148] рис. 3.2.

Вхідною точкою для роботи з даною системою є веб-орієнтований інтерфейс КІТ «Поліедр», який призначений як для використання простими користувачами, так і для адміністрування розробниками/експертами. Також даний інтерфейс управляє правами учасників процесу.

Режим використання надає можливість:

– застосовувати редукцію текстів. Для здійснення даного процесу необхідно надати системі вхідний ПМТ, який необхідно представити у вигляді онтології;

– здійснювати класифікацію текстів;

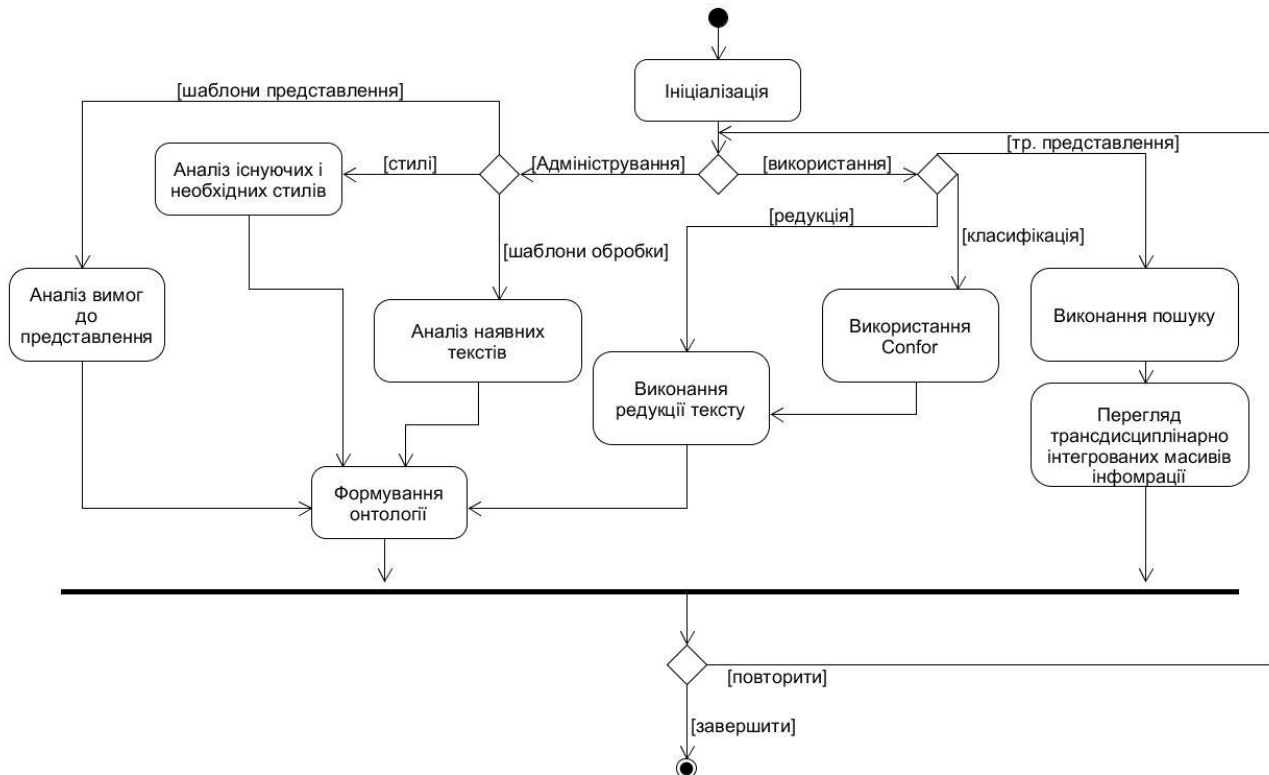


Рисунок 3.2 – Діаграма активності системи

– виконувати пошук інформації. Пошук інформації здійснюється в рамках індексних зон, кожна з яких представлена певним сервісом типу «Індексатор». Індексні зони можуть задаватись або в налаштуваннях додатку, або в налаштуваннях користувача. Можливий зовнішній, внутрішній та онтологічний пошук, призначений для відображення пошукових результатів у вигляді онтографу;

– користуватися режимом трансдисциплінарного представлення. В даному режимі користувач здійснює контекстне зв'язування інформаційних ресурсів, що в подальшому формує представлення, яке може бути відображене в якості трансдисциплінарної призми.

Режим адміністрування дозволяє:

– створювати шаблони представлення та обробки. Результати рекурсивної редукції замість редактора онтологій можуть завантажуватися в спеціальний адміністративний модуль, який дозволяє керувати базою правил;



– здійснювати аналіз існуючих та необхідних стилів. Експерти одночасно констатують наявність у мові різних функціональних стилів. Їх природа зумовлена об'єктивною реальністю. Проте, принципи виокремлення та класифікації стилів значною мірою залежать від погляду того чи іншого дослідника різних ПГ на способи вирішення поставлених задач.

### 3.2.3. Взаємодія системи

На рис 3.3 представлена діаграма взаємодії [147, 148].

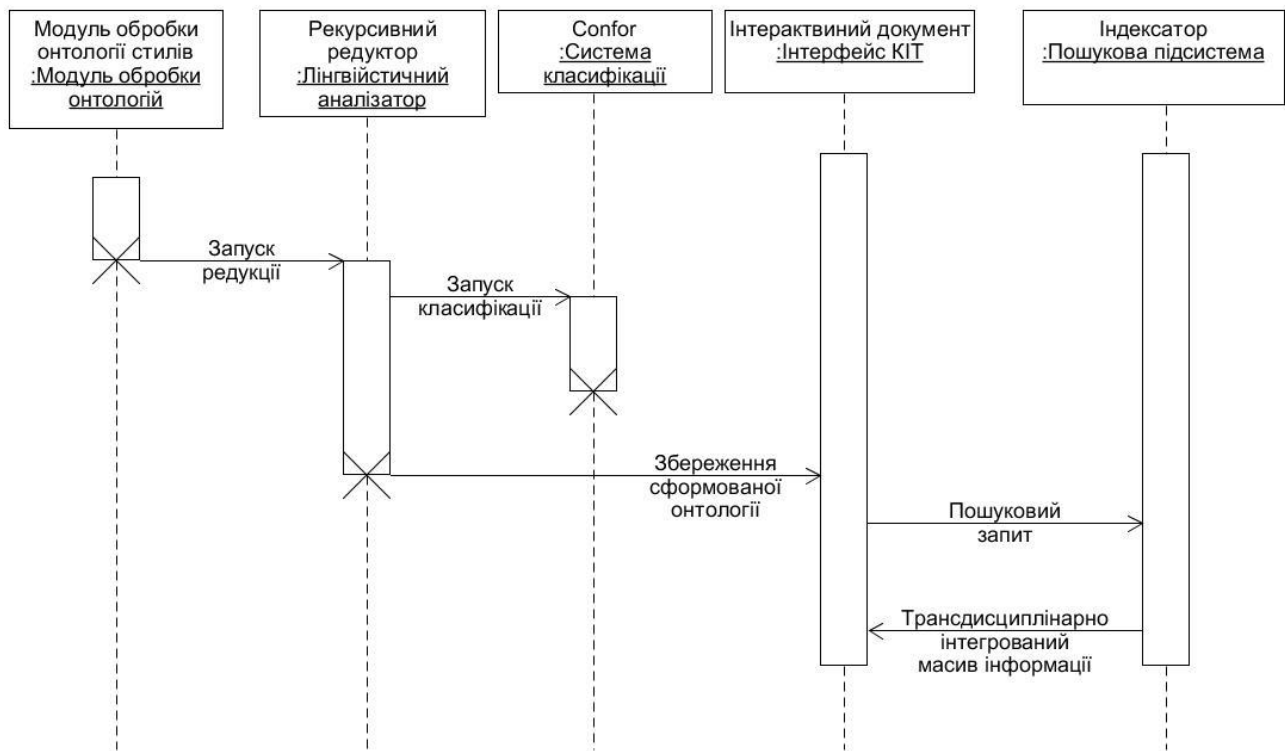


Рисунок 3.3 – Діаграма взаємодії системи

Дана діаграма описує основний процес системи трансдисциплінарного представлення IP різних стилів, а саме розширення функціональності та підвищення ефективності рекурсивної редукції [109] шляхом застосування дворівневої схеми обробки (з додатковою класифікацією вхідних текстів за стилем), що дозволяє значно підвищити точність вихідного результату і зменшити кількість помилок обробки.

Взаємодія підсистем відбувається в наступній послідовності:

- 1) У модулі обробки онтологій відбувається обробка онтології стилів;

- 2) Запускається редукція ПМТ до якої підключається класифікація тексту, що обробляється за стилем;
- 3) Сформована онтологія відображається інтерфейсом КІТ «Поліедр» у вигляді інтерактивного документа;
- 4) Елементи отриманого інтерактивного документа можуть виступати в якості пошукового запиту, виконавши який за допомогою індексації, система надає користувачеві трансдисциплінарно інтегрований масив інформації.

### 3.3. Архітектура системи

Для забезпечення надійної та швидкої роботи системи трансдисциплінарної консолідації ПМТ різних стилів використовується трирівнева архітектура, включаючи рівень представлення, рівень логіки та рівень даних (рис. 3.4).

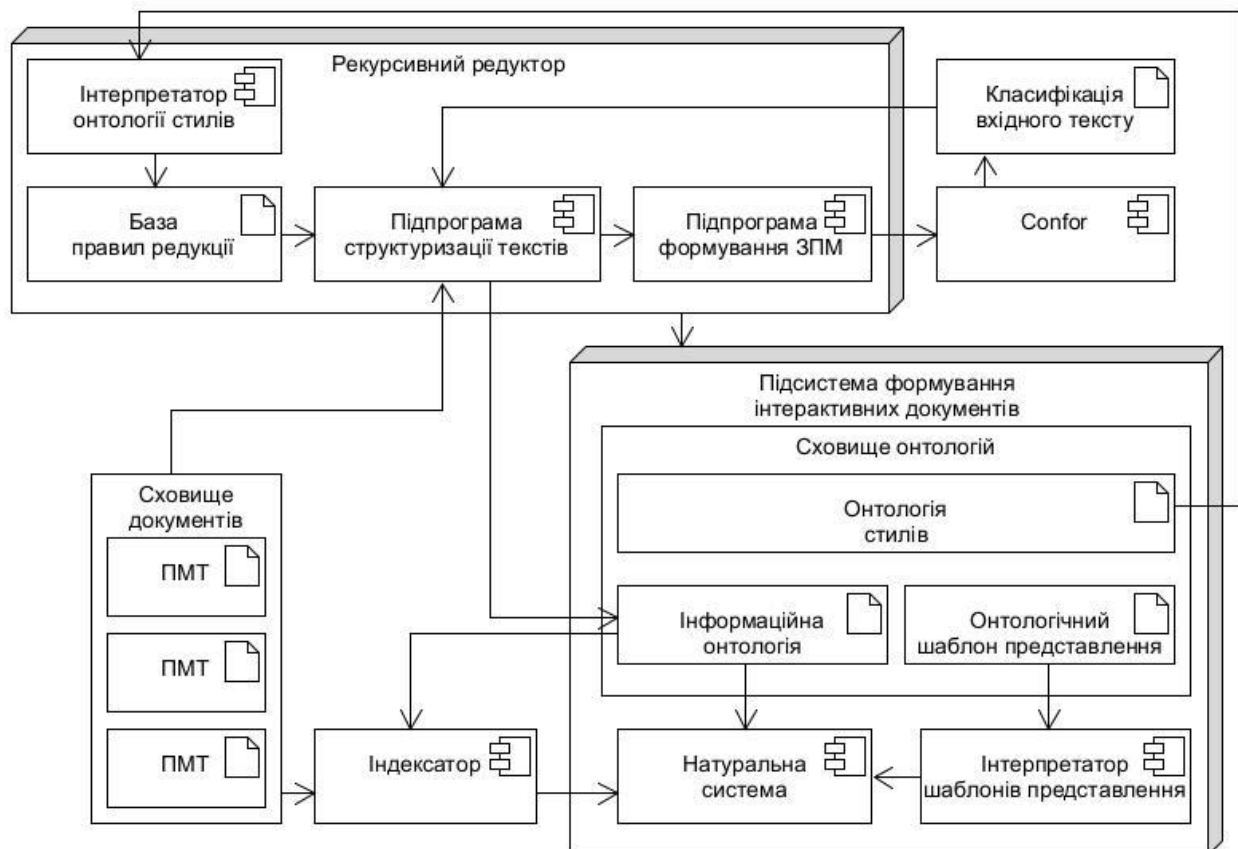


Рисунок 3.4 – Базова архітектура системи структурування документів різних стилів

Система складається з двох підсистем: підсистеми, що здійснює структурування текстів («Рекурсивний редуктор»), і підсистеми, що забезпечує користувацький інтерфейс («Підсистема формування інтерактивних документів»).

Система призначена для роботи з розподіленим масивом ПМТ, які можуть бути отримані з різних джерел. Центральним елементом системи є онтологія стилів, що міститься в сховищі онтологій в рамках підсистеми формування інтерактивних документів. Ця підсистема в тому числі забезпечує інтерфейс адміністратора для зміни онтології стилів.

При запуску підсистеми структуризації онтологія стилів витягується зі сховища, і за допомогою спеціалізованої підпрограми («інтерпретатор онтології стилів») перетворюється в підмножину бази правил редукції.

З урахуванням цих правил і здійснюється структуризація на першому етапі. Результати структуризації перетворюються в ЗПМ за допомогою відповідної підпрограми, і на їх основі проводиться класифікація вихідного тексту за допомогою класифікаційної системи «Confor». Отримана класифікація подається на вхід підпрограми структуризації, що дозволяє повторити редукцію з більшою точністю і отримати кінцевий результат – інформаційну онтологію, яка в подальшому зберігається у відповідній базі.

Підсистема формування інтерактивних документів містить дві основні складові: сховище онтологій і натуральну систему. Додатково використовується інтерпретатор онтологічних шаблонів представлення, що дозволяє модифікувати функціональність натуральної системи. Онтологічні шаблони представлення, що використовуються в роботі інтерпретатора, зберігаються в сховищі поруч з інформаційними онтологіями.

Додатково натуральна система дозволяє виконувати трансдисциплінарну інтеграцію інформації. Для цього використовується спеціалізований модуль індексації («Індексатор»). На вхід індексатора подаються як структуровані (інформаційні онтології), так і неструктуровані (ПМТ) документи, а результати його роботи відображаються за допомогою натуральної системи [84, 149].

При цьому кожна з підсистем має свої рівні логіки, даних та представлення. Для підсистеми формування інтерактивних документів рівень даних забезпечується сховищем онтологій, а рівні логіки та представлення – натуральною системою. Підсистема структуризації не передбачена для прямої взаємодії з користувачем і не має рівня представлення, рівень логіки забезпечується підпрограмою структуризації, а рівень даних – спеціалізованими підпрограмами зчитування і запису файлів (зокрема, інтерпретатором онтології стилів і підпрограмою формування ЗПМ).

Така архітектура дозволяє забезпечити максимально ефективну взаємодію різних частин системи між собою і системи в цілому – з користувачем.

### **3.4. Програмний засіб КІТ «Поліедр». Загальна структура**

Когнітивна ІТ-платформа (КІТ) «Поліедр» [150] призначена для підтримки процесів лінгвістично-семантичного аналізу великих обсягів просторово розподіленої неструктурованої інформації (Big Data), їх структуризації, встановлення контекстних зв'язків між документами, що обробляються, прогнозування та підтримки процесів раціонального вибору з наступним формуванням інформаційно-аналітичних WEB-орієнтованих рішень.

Засоби КІТ «Поліедр» забезпечують реалізацію наступних інформаційних процесів:

- Лінгвістично-семантичний аналіз мережевих інформаційних ресурсів, які мають значну кількість міждисциплінарних відношень, та створені на основі використання різних інформаційних технологій і стандартів;
- Трансдисциплінарну інтеграцію з іншими мережевими інформаційними системами та WEB-орієнтованими додатками;
- Таксономізацію наративів довільних документів та відображення їх структури, включаючи міжконтекстні зв'язки;
- Створення онтологічних інтерактивних документів;
- Виявлення латентної інформації в інформаційних ресурсах, що аналізуються;
- Глибинне та машинне навчання (Deep Learning, Machine Learning);

- Підтримку форматів і протоколів Semantic Web;
- Опрацювання великих даних (Big Data).

КІТ призначена для роботи в багатосерверному середовищі. В такому середовищі один сервер виконує основну роль (**сервер ядра**), на якому розміщуються наступні сутності:

- **«Переглядач онтологій»** являє собою *базовий додаток*, що дозволяє користувачам переглядати інформацію, наявну в онтологіях. Переглядач забезпечує інтерфейс для взаємодії з сервісами типу «Агрегована таблиця» та «Індексатор».
  - **«Пошукова призма»** – *додаток* на основі «Переглядача онтологій», що забезпечує спеціалізований інтерфейс для трансдисциплінарної консолідації інформаційних ресурсів з допомогою сервісу типу «Індексатор».
- **«Альтернатива»** – *базовий додаток*, призначений для багатокритеріального ранжування альтернатив, представлених об'єктами онтології.
- **«Редактор онтологій»** являє собою *базовий додаток*, що дозволяє користувачу створювати і редагувати онтології.
  - **«Редактор локальних файлів»** – *додаток* на основі «Редактора онтологій», оптимізований для роботи з файлами, що розміщені в рамках файлової системи робочого місця користувача.
  - **«Шаблонізований редактор»** – *додаток* на основі «Редактора онтологій», призначений для створення онтологій у відповідності шаблону (в формі спеціалізованої керуючої онтології).
- **«Онтологічне АРМ» (Онтологічне автоматизоване робоче місце)** – спеціалізований *базовий додаток*, призначений для здійснення бізнес-процесів, описаних з допомогою спеціалізованих керуючих онтологій (т. зв. онтологій процесу). Не має самостійної ролі і є базою для побудови

складних інформаційно-аналітичних систем, направлених на вирішення спеціалізованих задач.

Всі вказані вище типи додатків можуть додатково модифікуватись з допомогою онтологічних шаблонів представлення, в результаті чого створюються нові, більш спеціалізовані додатки, направлені на вирішення певної конкретної задачі.

Інші сервери багатосерверного середовища використовуються для розгортання і запуску **сервісів**. Як правило, користувач взаємодіє з сервісами через спеціалізовану підпрограму переглядача онтологій, хоча бувають сервіси, що надають власний інтерфейс. Зовнішніх серверів може бути довільна кількість. Основні типи сервісів, що можуть розміщуватись на зовнішніх серверах:

- **Індексатор** – *сервіс*, призначений для повнотекстового пошуку інформації у великих масивах структурованих і неструктурованих документів;
- **Агрегована таблиця** – *сервіс*, призначений для зберігання великих масивів структурованої інформації;
- **Конспект** – *сервіс*, призначений для структуризації неструктурованих документів шляхом виділення з них термінів (має власний інтерфейс).

Робоче місце користувача також є складовою середовища і використовується для запуску **локальних сервісів**:

- **Рекурсивний редуктор** – *локальний сервіс*, призначений для структуризації слабко- і неструктурованої інформації з допомогою спеціалізованих правил, заданих в форматі  $\lambda$ -виразів;

**Підсистема природномовних лінгвоінформаційних операцій** – локальний сервіс, призначений для порівняння природномовних текстів.

### 3.5. Висновки за розділом 3

Розроблено інформаційну модель програмної системи, яка реалізує трансдисциплінарне представлення різностильових ІР, що включає опис модулів: інтерпретації онтології стилів; формування ЗПМ; класифікації вхідного тексту за стилем; інтерпретації шаблонів представлення.

Розроблено модель поведінки системи, що описує: варіанти використання системи; активність системи; взаємодію системи.

На основі інформаційної моделі та моделі поведінки системи трансдисциплінарного представлення різностильових ІР розроблено її архітектуру, яка визначає структуру, компонентний склад і функціональність розроблених програмних засобів.

Результати досліджень, наведені в третьому розділі, опубліковано в роботах [83, 149]

## **РОЗДІЛ 4. ПРОГРАМНЕ СЕРЕДОВИЩЕ ТРАНСДИСЦИПЛІНАРНОГО ПРЕДСТАВЛЕННЯ ПРИРОДНОМОВНИХ ВИДАНЬ**

Система автоматизованої обробки різностильових документів являє собою *спеціалізований додаток* КІТ «Поліедр».

Програмна платформа «Поліедр» дозволяє створювати додатки з допомогою спеціалізованих онтологій (керуючих та інформаційних), що виступають в якості онтологічних шаблонів представлення. Можна виділити наступні сутності, які є основою функціонування системи:

1. Модуль класифікації та структуризації – програмний модуль, що є розширенням модулю «Рекурсивний редуктор» із стандартної бібліотеки модулів КІТ.
  - 1.1. Онтологічний класифікатор стилів – керуюча онтологія, що формалізує використовувану системою підмножину стилів і задає правила обробки відповідних документів.
  - 1.2. Підсистема інтеграції з модулем «Confor» стандартної бібліотеки КІТ – призначена для автоматизації взаємодії з модулем шляхом формування файлів в форматі, сумісному з Confor, і зчитування результатів його роботи.
2. Модуль відображення і трансдисциплінарної консолідації інформації – розширення стандартного модулю перегляду онтологій КІТ

Онтологічний шаблон представлення модулю відображення і трансдисциплінарної консолідації – дозволяє модифікувати стандартний переглядач КІТ, надаючи йому додаткові функції щодо більш тісної інтеграції з сервісом «Індексатор» з стандартної бібліотеки модулів КІТ.

### **4.1. Онтологія стилів та її формування за допомогою стандартних засобів КІТ «Поліедр»**

Першим етапом створення спеціалізованого додатку «Система автоматизованої обробки різностильових документів» є створення керуючої онтології «Онтологічний класифікатор стилів» за допомогою наступних функцій:



### 4.1.1. Функція створення об'єкту

Функція створення об'єкту дозволяє додавати до онтології стилів нові категорії стилів і власне стилі. Здійснюється з допомогою ієрархічного фільтру (рис. 4.1 (2)) або режиму редагування онтографу (рис. 4.1 (3)).

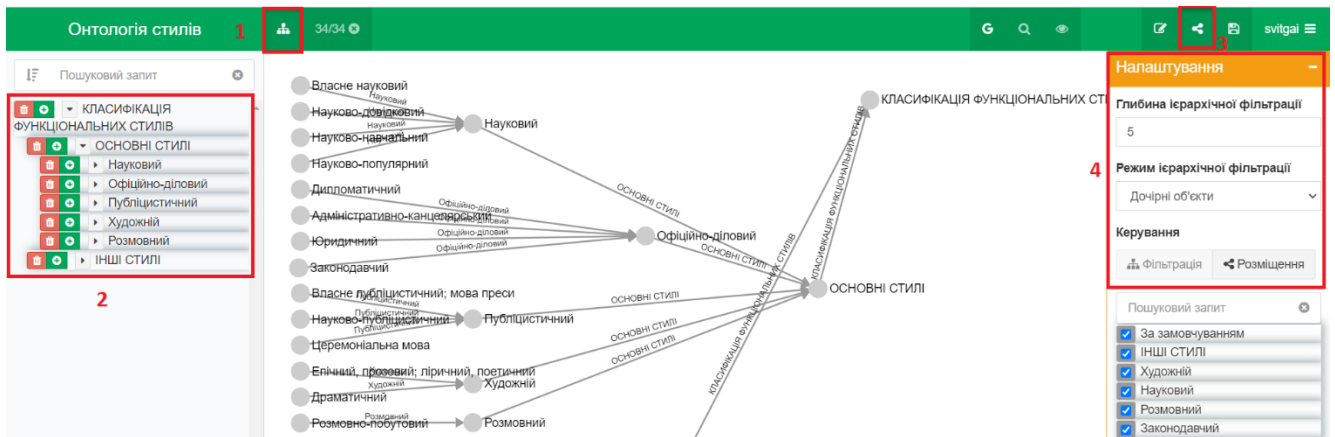


Рисунок 4.1 – Створення об'єктів онтології

Ієрархічний фільтр відображає ієрархію об'єктів і працює наступним чином:

- 1) Фільтр відкривається за допомогою кнопки виклику/приховання ієрархічного фільтру (рис. 4.1(1));
- 2) При відкритому Редакторі без вибраної онтології (або повністю видалених об'єктах) кнопка «Створити об'єкт» (рис. 4.2(1)) дозволяє створити кореневий об'єкт онтології;
- 3) Біля кожного об'єкта, наявного в онтології, зліва від стрілки відображається дві кнопки – видалення поточного об'єкту (рис. 4.2(2)) і додавання дочірнього об'єкту (рис. 4.2(3)).

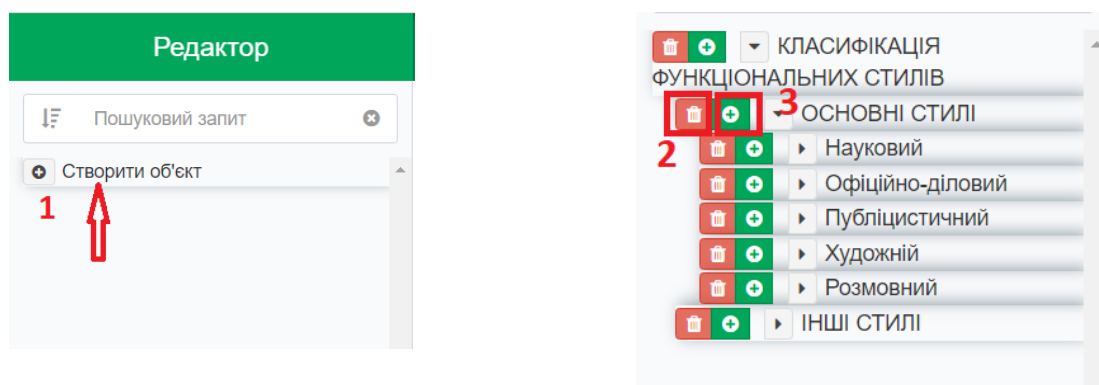


Рисунок 4.2 – Управління ієрархічним фільтром

Онтологія, сформована виключно з допомогою ієрархічного фільтру, буде деревовидною. В фільтра немає можливості надавати об'єкту кілька батьківських об'єктів, а при спробі створити об'єкт з назвою, яка вже присутня в онтології, буде здійснене автоматичне перейменування.

Оскільки ієрархічний фільтр не дозволяє вибрати позицію об'єкту в онтографі, то така позиція буде обчислена автоматично на основі координат батьківського об'єкту (рис. 4.3). При цьому поточна онтологія буде відмічена як така, що містить невалідні координати.

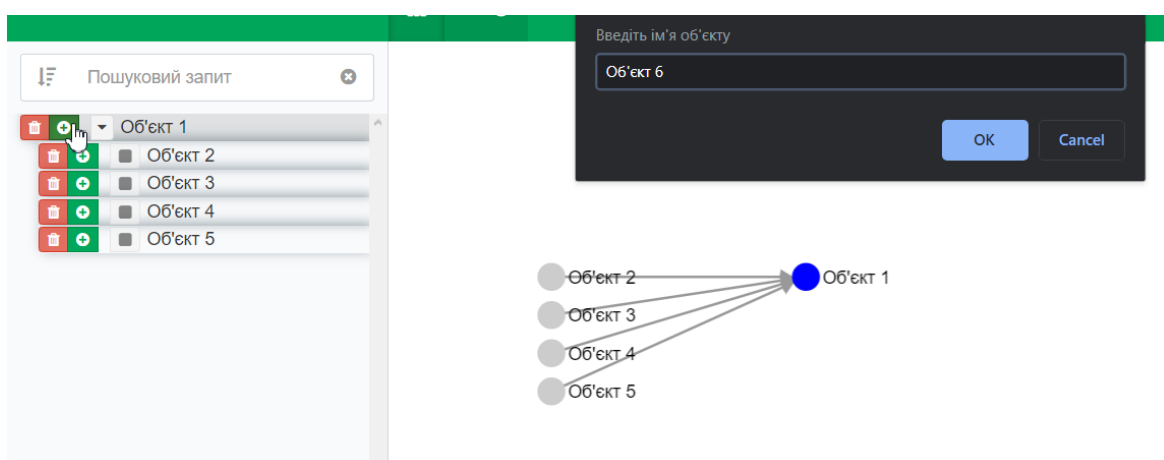


Рисунок 4.3 – Розміщення створених ієрархічним фільтром об'єктів

Інтерфейс режиму редагування онтографу (рис. 4.1 (4)) викликається кнопкою, як показано на рис. 4.1(3) і містить кнопку «Розміщення», яка запускає цей процес.

Створення об'єктів здійснюється за допомогою «гарячої клавіші»: Ctrl+клік на пустому місці в робочій області створює об'єкт в даній позиції. При цьому надається запит на введення імені об'єкту. Ім'я має бути унікальним, інакше буде здійснене перейменування.

#### 4.1.2. Функція створення зв'язків

Функція створення зв'язків дозволяє будувати структуру онтології. В режимі редагування онтографу зв'язки створюються наступним чином:

Ctrl+клік на існуючому об'єкті відмічає його як початковий для створення зв'язку. Ctrl+клік на тому ж самому об'єкті відміняє створення зв'язку, а на іншому

– створює між ними зв'язок. Всі етапи даного процесу супроводжуються повідомленнями.

Додатково може використовуватись більш простий спосіб побудови структури – з допомогою ієрархічного фільтру. Кнопки ієрархічного фільтру дозволяють створювати дочірні об'єкти до поточного, тобто власне новий об'єкт і зв'язок між ним і поточним.

Обмеженням ієрархічного фільтру є неможливість створення більш ніж одного кореневого об'єкту і присвоєння об'єкта кільком категоріям. При необхідності здійснення такої дії використовується режим редагування онтографу і всі потрібні виправлення вносяться з його допомогою.

Структура онтології стилів має чотирирівневу ієрархію об'єктів (рис. 4.4), де вершини 1, 2 і 3 рівня – це категорії, що не мають атрибутів, а вершини 4 рівня – це приклади конкретних документів певного стилю. Вершини 4 рівня мають атрибути.

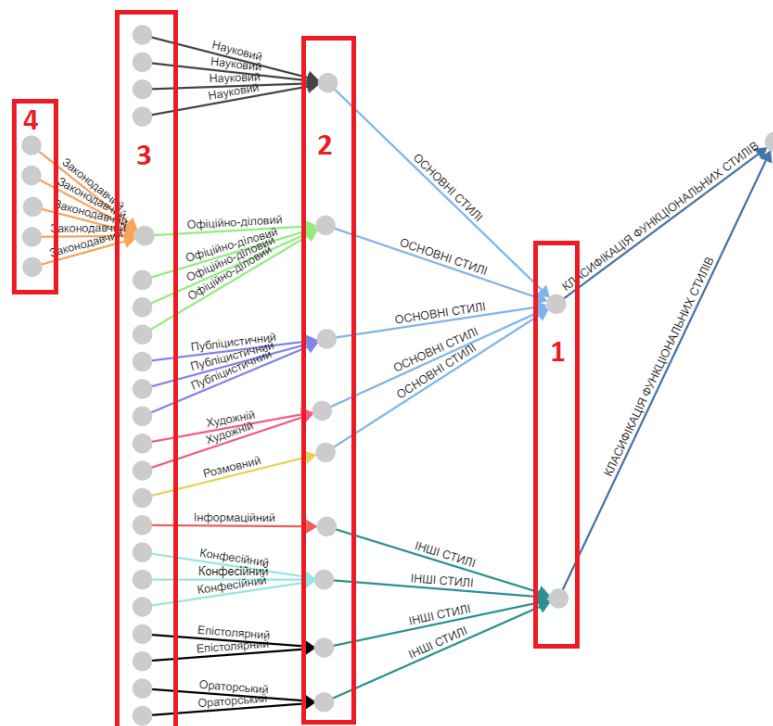


Рисунок 4.4 – Структура онтології стилів

### 4.1.3. Функція видалення об'єктів і зв'язків

Функція видалення об'єктів і зв'язків дозволяє редагувати структуру онтології.

Помилково введені об'єкти чи зв'язки видаляються кліком з натиснутою клавішею Alt.

При необхідності виділення кількох об'єктів, використовується клавіша Shift.

Клавіша Delete використовується для видалення виділених об'єктів. Будь-яке видалення супроводжується запитом підтвердження.

Листовий об'єкт (без дочірніх об'єктів) можна видалити кнопкою в ієрархічному фільтрі, однак для об'єкту-категорії (з дочірніми об'єктами) дана кнопка видає повідомлення про помилку. Об'єкт-категорія, за необхідності, видаляється в режимі редагування онтографу. При цьому якщо якісь дочірні об'єкти відносяться тільки до одної категорії, то вони стануть корневими (рис. 4.5).

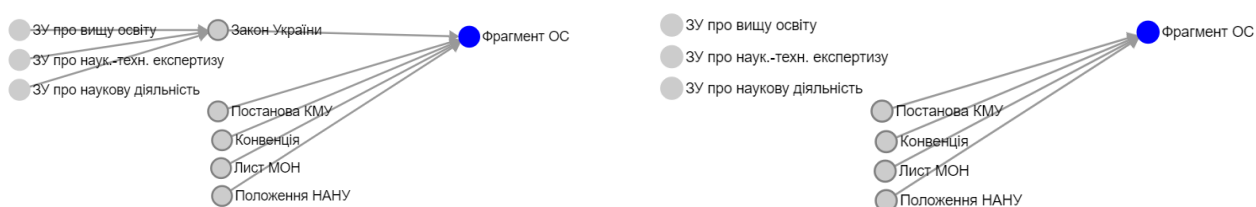


Рисунок 4.5 – Результат видалення категорії

При необхідності видалення всього піддерева, використовується наступний алгоритм:

1) В режимі редагування онтографу вибирається кореневий об'єкт призначений для видалення піддерева в ієрархічному фільтрі (або виділяється кліком і натисканням кнопки «Фільтрація» у віджеті) (рис. 4.6, а);

2) У віджеті вибирається режим ієрархічної фільтрації «Дочірні об'єкти» і встановлюється необхідна глибина ієрархічної фільтрації;

3) Виділяються всі об'єкти комбінацією Ctrl+A (рис. 4.6, б), та видаляється обрана підмножина об'єктів кнопкою Delete. Результат видалення піддерева об'єктів зображено на рис. 4.8.

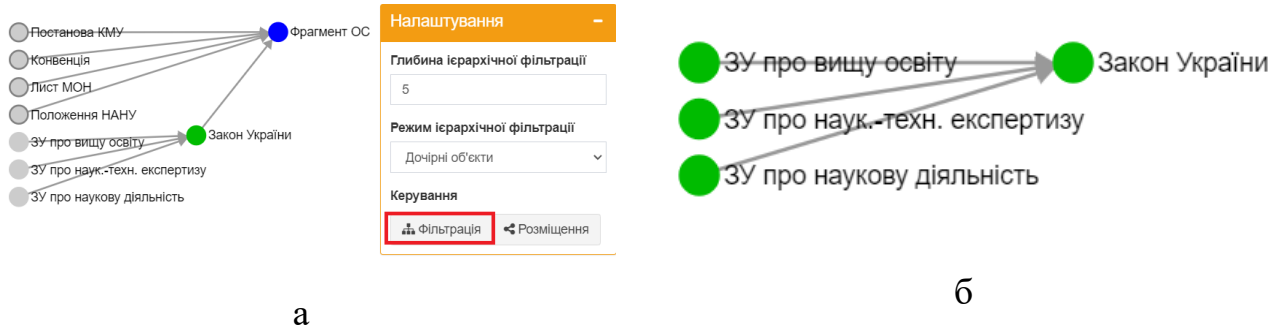


Рисунок 4.6 – Виділення піддерева об'єктів



Рисунок 4.7 – Результат видалення піддерева об'єктів

#### 4.1.4. Функція редагування атрибутів

Функція редагування атрибутів дозволяє створювати, редагувати та видаляти атрибути стилів.

Для редагування атрибутів використовуються:

- Режим редагування атрибутів;
- Вікно редагування атрибутів, що викликається подвійним кліком на об'єкті в режимі редагування онтографу.

Дані компоненти майже ідентичні за своїми функціями і структурою. Загальний вигляд режиму редагування атрибутів показано на рис. 4.8.

Інтерфейс для редагування атрибутів ділиться на дві частини: заголовок (призначений для введення загальної інформації про об'єкт, такої, як назва чи колір) і основну область (призначену для редагування списку атрибутів. Заголовок містить (рис. 4.9):

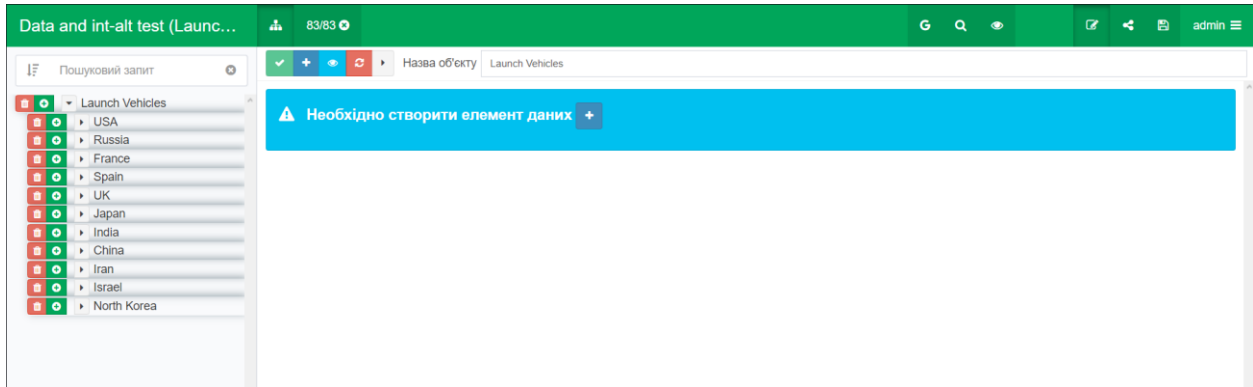


Рисунок 4.8 – Режим редагування атрибутів

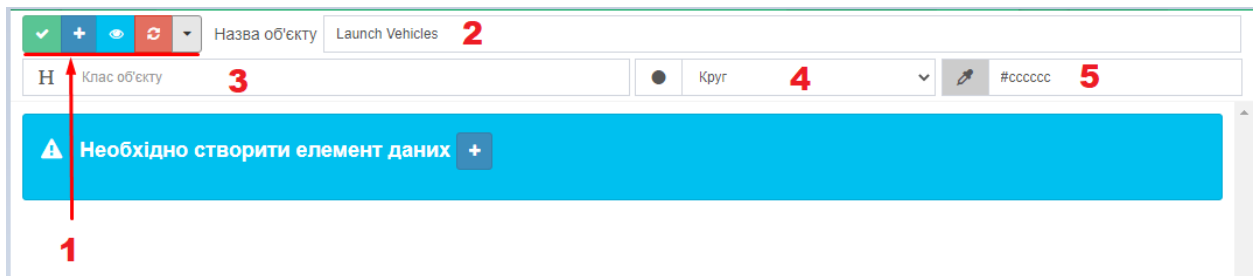


Рисунок 4.9 – Заголовок режиму редагування атрибутів

- 1) Блок керуючих кнопок (1) (зліва направо):
  - a. Кнопка збереження змін – активується при внесенні будь-яких змін в інтерфейсі і застосовує ці зміни до об'єкту;
  - b. Кнопка додавання атрибуту (елементу даних) – додає блок атрибуту на першу позицію в списку. Якщо об'єкт не має атрибутів, то в основній області буде відображене повідомлення з такою ж кнопкою;
  - c. Попередній перегляд – запускає вікно детального перегляду атрибутів із поточним вмістом робочої області. При цьому зміни до об'єкта не застосовуються;
  - d. Кнопка скидання змін – відмінює всі зроблені в інтерфейсі зміни і виводить в робочу область поточні дані об'єкту;
  - e. Кнопка показу додаткових налаштувань – показує/приховує другий рядок заголовку, що містить другорядні метадані (клас, форму і колір).
- 2) Назва об'єкту (2) – поле, що дозволяє перейменувати об'єкт.
- 3) Клас (3) – спеціалізоване значення, що застосовується деякими видами додатків. В загальному випадку не використовується.

- 4) Форма (4) – задає форму об'єкту для режимів представлень онтографу і ГІС-додатку. Основними є круг і квадрат, додатково є варіації з вкладеними в них картинками.
- 5) Колір (5) – колір об'єкту для режимів відображення онтографу і ГІС-додатку.

Список атрибутів містить блоки, кожен з яких представляє певний атрибут.

Блок містить (рис. 4.10):

- 1) Заголовок – верхню частину, з загальними керуючими елементами, а саме:
  - a. Кнопка «Додати блок після поточного» (1);
  - b. Кнопка «Видалити поточний блок» (2);
  - c. Перемикач типу атрибуту (3);
  - d. Поле вводу назви (4).
- 2) Робочу область – основну частину, вміст якої залежить від вибраного типу атрибуту.

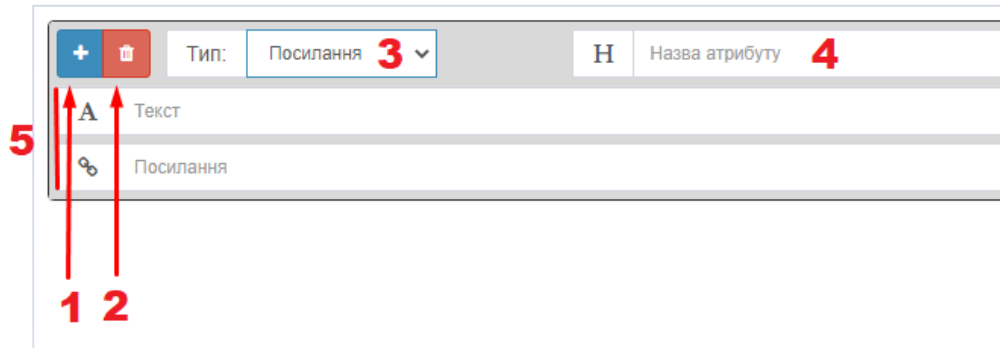


Рисунок 4.10 – Загальний вигляд блоку атрибуту

Існують наступні типи атрибутів:

- 1) Текстові дані;
- 2) Дані-гіперпосилання;
- 3) Зображення;
- 4) Відео з YouTube;
- 5) Форматований текст;
- 6) Складні типи даних (групи і таблиці);
- 7) Геопросторові дані.

Для онтології стилів використовуються таблиці даних, що являють собою певну ієрархію атрибутів. Таблиця даних являє собою власне двовимірну таблицю, призначену для відображення великого списку однакових сутностей. Робоча область містить елементи для вибору розміру таблиці (рис. 4.11(1)), за допомогою яких встановлюється необхідна кількість рядків і стовпців таблиці, а також кнопку показу/приховування таблиці (рис. 4.11(2)), призначену для більш компактного відображення списку атрибутів у разі їх великої кількості.

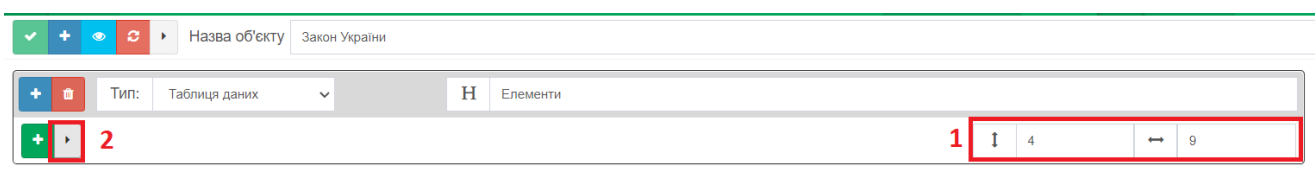


Рисунок 4.11 – Формування атрибутів типу "Таблиця даних"

Для формування таблиці даних, що відображає унікальність текстової розмітки конкретних законодавчих актів, дані документи аналізуються за допомогою стандартних засобів Microsoft Word (рис. 4.12).

Таким чином визначаються унікальні елементи документу. В даному випадку це: назва типу документу (1), назва документу (2), назва розділу (3), назва статті (4). А також параметри розмітки цих елементів: розмір шрифту (5), жирний (6), курсив (7), назва шрифту (8). Для інших документів можна використовувати також підкреслення, вирівнювання, тощо. Для підвищення точності обробки при можливості застосовуються елементи, що явно вказують на фрагменти тексту, які необхідно виділити. Наприклад, в документі що аналізується, перед назвою розділу завжди вказано слово «РОЗДІЛ» (3); перед назвою статті – «Стаття» (4).

Для внесення визначених даних в онтологію стилів, у вікні формування атрибутів (для об'єкту «Закон України») створюється таблиця з чотирьох рядків і шести стовпчиків (рис. 4.13). Де в першому стовпчику вказано елементи документу, а в першому рядку – елементи розмітки. На перетині комірок вказано значення кожного елементу розмітки для кожного елементу документу.



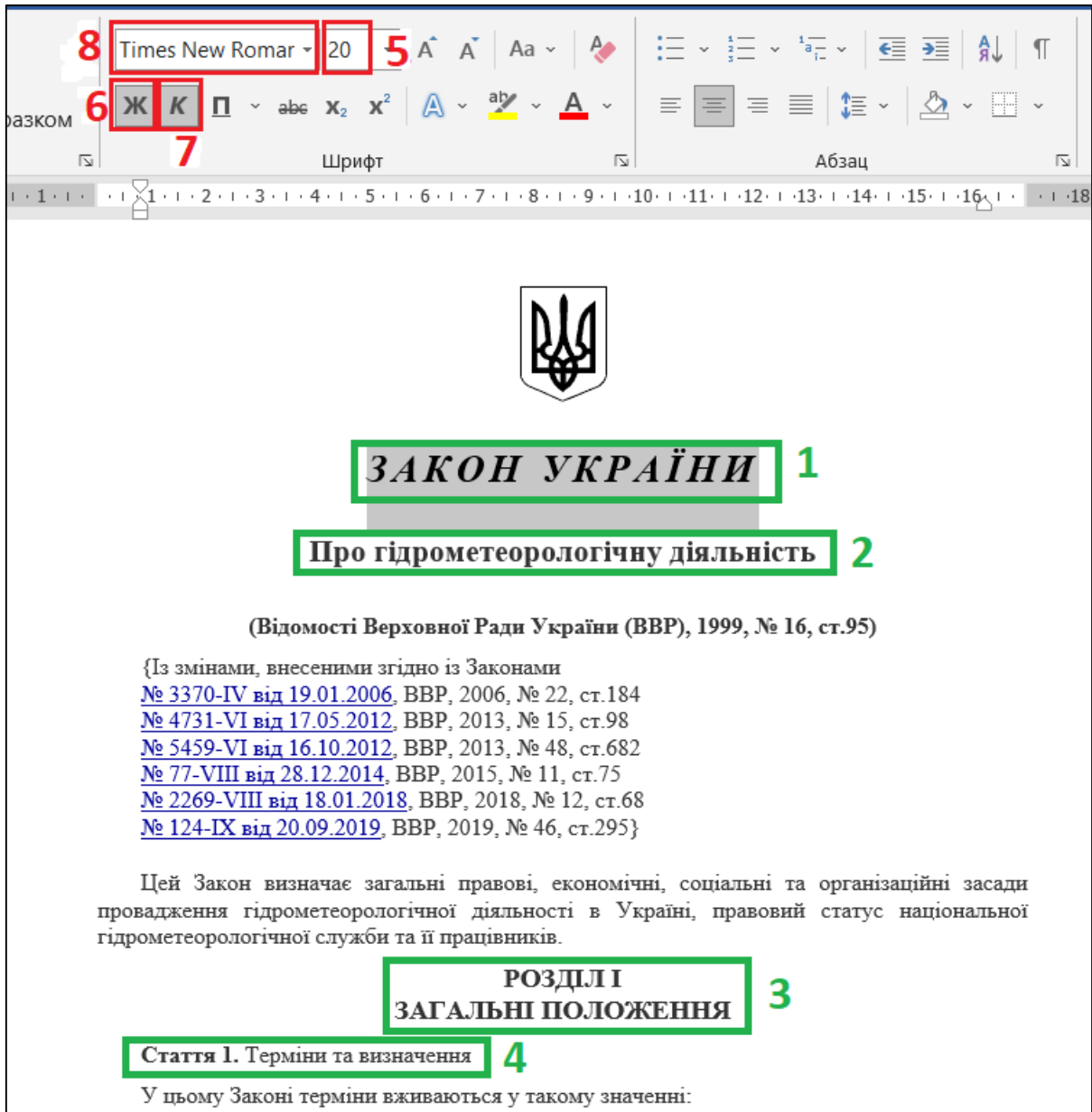


Рисунок 4.12 – Приклад законодавчого акту

Назва	Розмір шрифту	Жирний	Курсив	Шрифт	Початок
Назва типу документу	20	Так	Так	Times New Roman	
Назва документу	16	Так		Times New Roman	
Назва розділу	14	Так		Times New Roman	РОЗДІЛ
Назва статті	12	Так		Times New Roman	Стаття

Рисунок 4.13 – Таблиця даних для об'єкту "Закон України"

На практиці, процес створення таблиць даних для великої кількості об'єктів є достатньо трудомістким через необхідність постійного порівняння і узгодження кількості і назв однотипних елементів для різних об'єктів. Тому, для створення таблиць даних доцільно використовувати імпорт з XLSX. Цей спосіб введення даних дозволяє одночасно бачити всі таблиці у зведеному вигляді.

#### 4.1.5. Функція зчитування атрибутів з XLSX-файлів

Функція зчитування атрибутів з XLSX-файлів надає альтернативний спосіб введення атрибутів стилів.

Формування таблиці атрибутів в Excel показано на рис. 4.14.

1	2	3	4	5	6	7	8	9	10			
	1	подерproperties	2	Назва	Розмір шрифту	Жирний	Курсив	Підкреслений	5	Шрифт	Початок	Регулярний вираз
4	Закон України	Назва типу документу	20	Так	Так				Times New Roman			
5	Закон України	Назва документу	16	Так					Times New Roman			
6	Закон України	Назва розділу	14	Так					Times New Roman	РОЗДІЛ		
7	Закон України	Назва статті	12	Так					Times New Roman	Стаття		
8	Положення класифікаційної діяльності	Назва типу документу	26	Так					Times New Roman			
9	Положення класифікаційної діяльності	Назва документу	24	Так					Times New Roman			
10	Положення класифікаційної діяльності	Назва розділу	12	Так					Times New Roman			
11	Положення класифікаційної діяльності	Назва підрозділу	11	Так					Times New Roman			
12	Положення класифікаційної діяльності	Терміни	11		Так				Times New Roman			
13	Кодекс	Назва документу	18	Так					Times New Roman			
14	Кодекс	Назва розділу	14	Так					Times New Roman	РОЗДІЛ		
15	Кодекс	Назва глави	14	Так					Times New Roman	ГЛАВА		
16	Кодекс	Назва статті	12	Так					Times New Roman	Стаття		
17	Кодекс	Відмітка про редагування	12		Так				Times New Roman			{.*}

Рисунок 4.14 – Фрагмент таблиці атрибутів poderproperties

Загальна структура таблиці атрибутів:

- 1) В комірці A1 розташовується ключове слово poderproperties (1), інакше таблиця буде оброблена як таблиця структури;
- 2) Таблиця може мати один або кілька рядків заголовків (багаторівневі заголовки). Кількість рядків в заголовку визначається кількістю комірок, з якими об'єднана A1 (наприклад, якщо A1 об'єднана з A2, це означає два рядки заголовків);
- 3) Інші комірки заголовків містять назви атрибутів (2);
- 4) Нижче знаходяться рядки, що відповідають об'єктам. В стовпчику A розміщується назва об'єкту, якому належать атрибути (3);

- 5) Об'єкту також можуть відповідати кілька рядків (4);
- б) Праворуч від назви об'єкту знаходяться значення його атрибутів;

Назви або значення атрибутів можуть мати модифікатор типу вкінці (`\text`, `\link`, `\image` `\table` та ін. (5)). Це дозволяє явно вказати тип атрибута. Якщо модифікатор знаходиться в заголовку, то він застосовується до всіх значень в стовпчику. Якщо модифікатор знаходиться в значенні атрибуту, то він застосовується тільки для даного значення (вищий пріоритет ніж в модифікатора в заголовку).

#### 4.1.6. Функція збереження в бібліотеку

Функція збереження в бібліотеку дозволяє створювати різні версії онтології або різні онтології під різні набори стилів.

Для збереження створених онтологій використовується кнопка збереження, розташована в заголовку вікна Редактора онтологій (рис. 4.15). Вона відображається в двох варіантах: «Зберегти як» і «Зберегти».



Рисунок 4.15 – Кнопка збереження онтологій

Кнопка «Зберегти як» дозволяє завантажити поточну онтологію в особисту бібліотеку користувача. При цьому користувачу видається стандартний запит на введення імені. Кнопка відображається тільки якщо поточний користувач авторизований.

Кнопка «Зберегти» відображається на місці кнопки «Зберегти як» якщо поточний користувач не тільки авторизований, а й має права на запис поточного файлу (в публічній або особистій бібліотеці).

Редактор онтологій надає можливість редагування списку онтологій в бібліотеці, зокрема:

- 1) Створювати і видаляти елементи бібліотеки (онтології і каталоги онтологій);
- 2) Редагувати елементи (мінати назву та інші атрибути);
- 3) Переміщувати елементи з одного каталогу в інший.

Загальний вигляд бібліотеки показано на рис. 4.16.

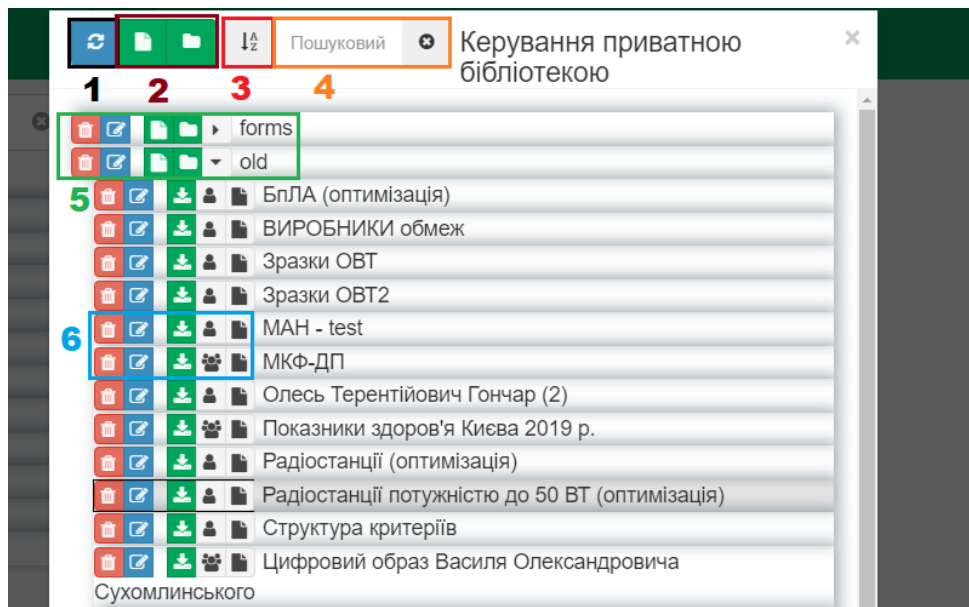


Рисунок 4.16 – Загальний вигляд керуючої бібліотеки

Основні керуючі елементи, що містяться в заголовку:

- 1) Кнопка оновлення бібліотеки (1);
- 2) Блок додавання елементів в корінь бібліотеки (2) – викликає форми створення елементів (рис. 4.18, рис. 4.19), а також служить областю, на яку можна перетягувати елементи бібліотеки. Блок відображається тільки якщо користувачу надано повний доступ до бібліотеки – такий доступ завжди є для особистої, а для публічної надається адміністратором;
- 3) Кнопка сортування (3) – дозволяє вибирати режим сортування:
  - а. Сортування в лексикографічному порядку – застосовується за замовчуванням (рис. 4.16).
  - б. Сортування за датою створення (рис. 4.17). В цьому режимі додатково відображається власне дата створення – справа від назви.

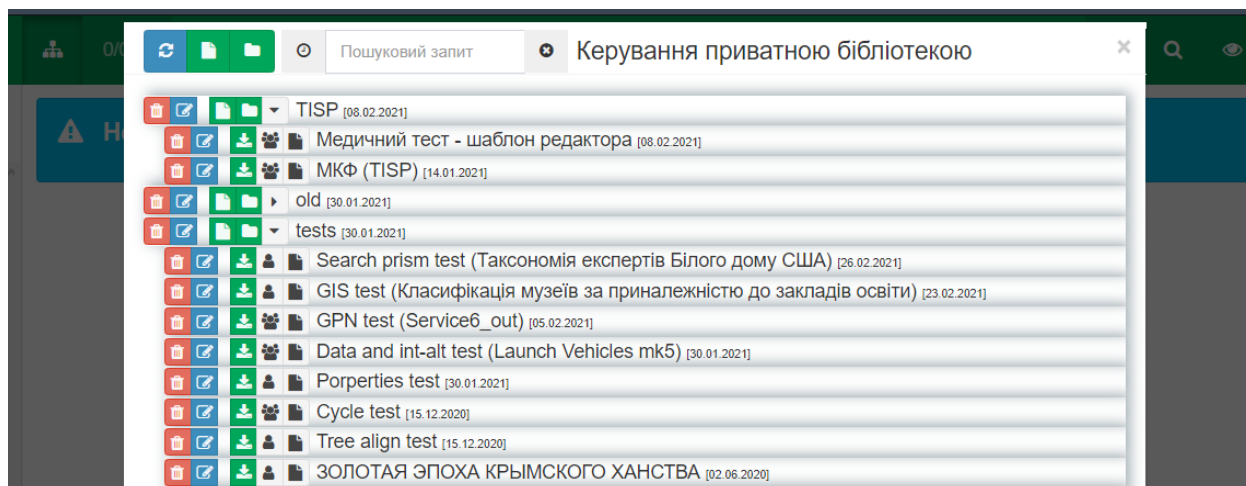


Рисунок 4.17 – Сортування бібліотеки за датою створення

#### 4) Блок фільтрації за назвою (4).

Кожен елемент бібліотеки має свій власний блок керування. Для каталогів онтологій (5) блок містить (зліва направо):

1) Кнопку видалення – для непустих каталогів видалення може бути заборонене адміністратором, в такому випадку спочатку необхідно буде видалити всі дочірні елементи.

2) Кнопку редагування – викликає форму редагування каталогу, що ідентична показаній на рис. 4.18 (за виключенням поля «Батьківський каталог»).

3) Кнопку додавання онтології – дозволяє додати онтологію в поточний каталог (викликає показану на рис. 4.19 форму).

4) Кнопку додавання каталогу – дозволяє додати дочірній каталог в поточний (викликає показану на рис. 4.18 форму).

5) Стрілку, що дозволяє показувати/приховувати дочірні елементи каталогу.

6) Назву каталогу, що діє аналогічно до стрілки.

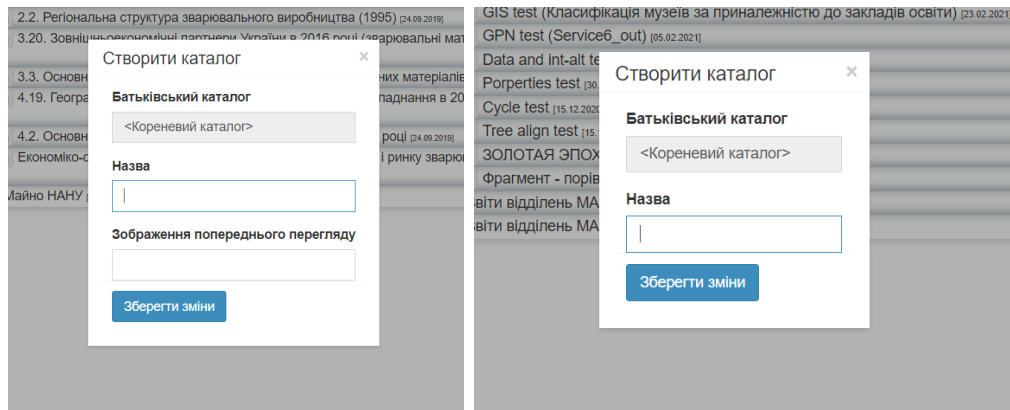


Рисунок 4.18 – Форми створення публічного (зліва) і особистого (справа) каталогу

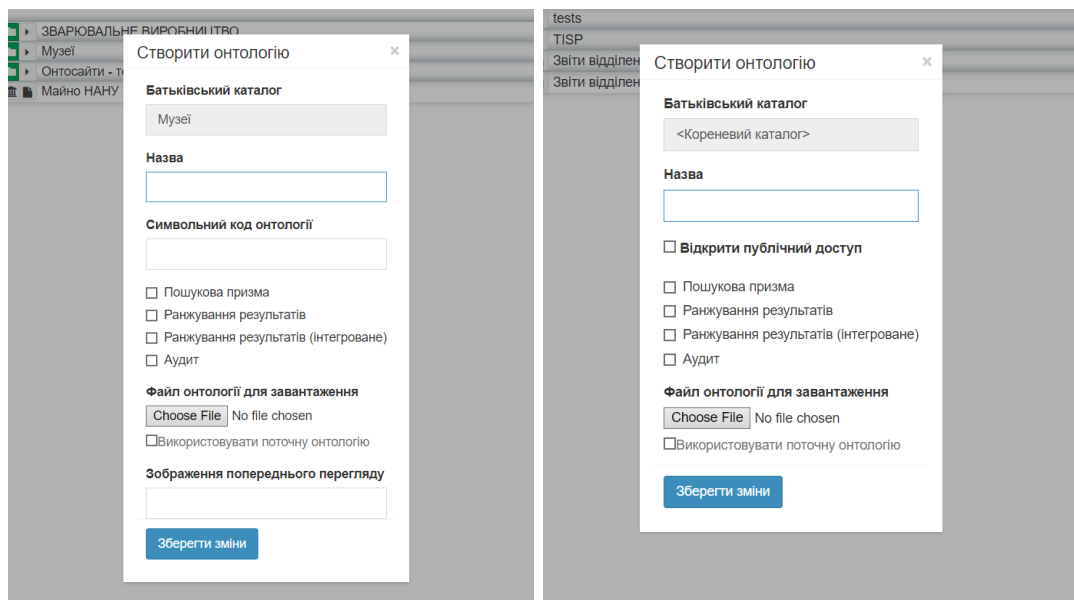


Рисунок 4.19 – Форми створення публічної (зліва) і особистої (справа) онтології

Для елементів типу «Онтологія» блок містить (зліва направо):

- 1) Кнопку видалення.
- 2) Кнопку редагування – викликає форму редагування онтології (рис. 4.20).
- 3) Кнопку збереження онтології на локальний диск у вигляді файлу.
- 4) Кнопку відкриття онтології в сконфігурованому «Переглядачі онтологій за замовчуванням) – кнопка також містить маркер типу файлу.
- 5) Кнопку відкриття онтології поточним додатком.
- 6) Назву, натиск на яку також відкриває онтологію в поточному додатку.

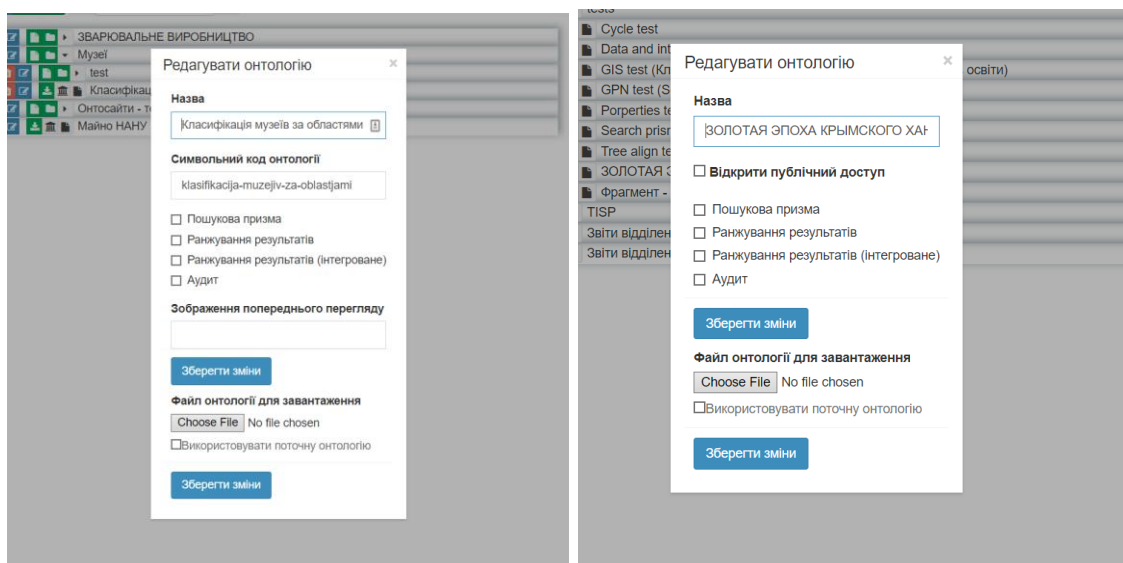


Рисунок 4.20 – Форми редагування публічної (зліва) і особистої (справа) онтології

#### 4.1.7. Функція роботи з буфером обміну

Функція роботи з буфером обміну дозволяє створювати нові онтології шляхом об'єднання елементів існуючих.

Буфер обміну задає стандартизований інтерфейс для обміну об'єктами між онтологіями, який також слугує для взаємодії деяких модулів. Дану функцію підтримують всі представлення онтографу (режим перегляду онтографу, онтологічний пошук), однак використовувати її може тільки експерт ПГ з доступом до Редактора онтологій.

Для початку обміну, у відкритій онтології, що служить джерелом даних виділяється потрібна підмножина об'єктів (або всі об'єкти – комбінацією Ctrl+A).

Ctrl+C ініціює копіювання в буфер обміну.

Через обмеження безпеки копіювання здійснюється не напряму, а з допомогою форми (рис. 4.21) (для сучасних браузерів достатньо натиснути Enter або кнопку «Скопіювати...» (1), вікно буде автоматично закрито. При цьому може бути показане попередження безпеки (2). Якщо даний механізм не спрацює в поточному браузері, вміст текстового поля копіюється вручну (3), після чого закривається вікно.

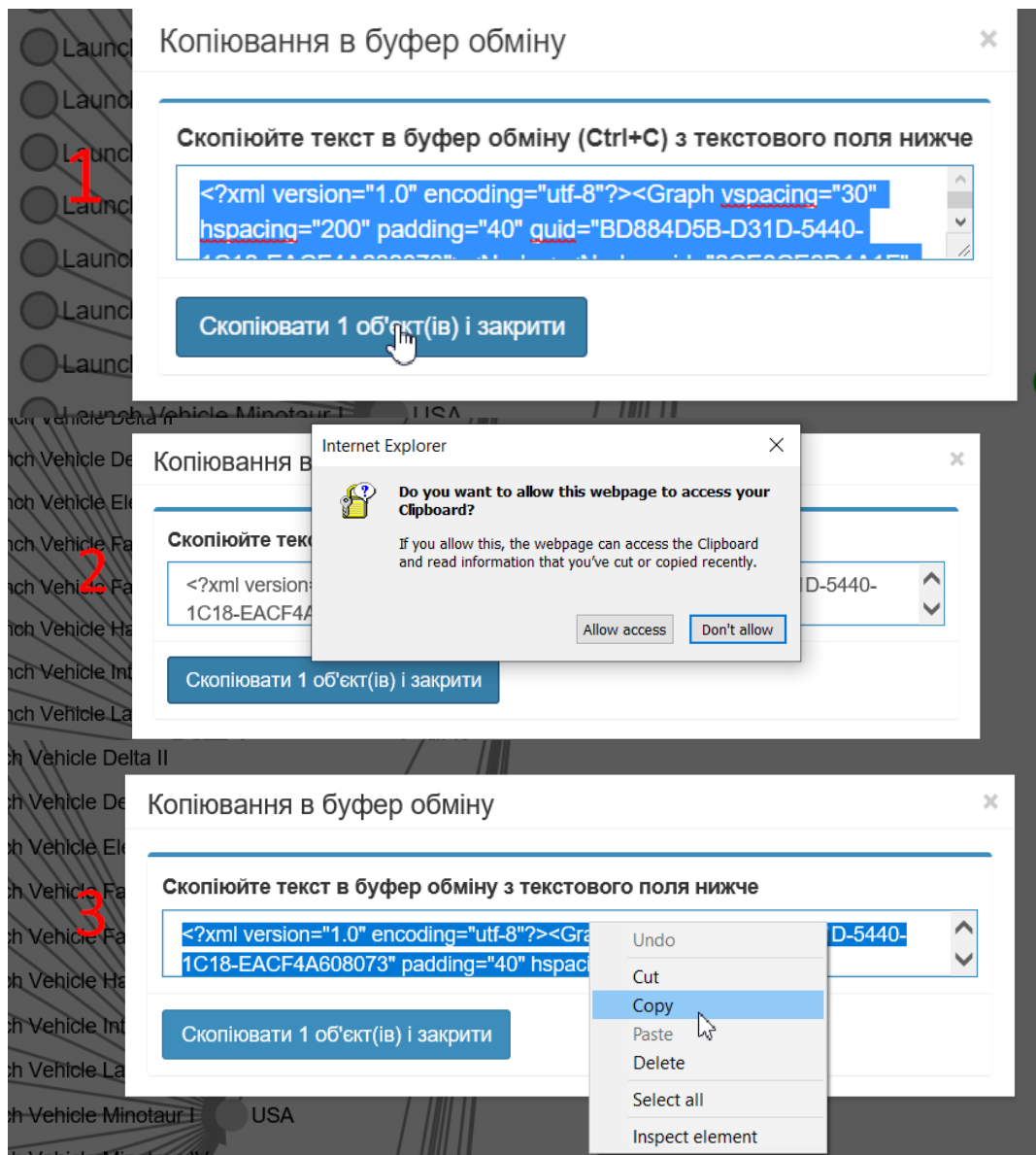


Рисунок 4.21 – Вікно копіювання

Ctrl+V працює тільки в режимі редагування онтографу, і вставляє наявний в буфері обміну фрагмент XML-онтології в поточну робочу область. При цьому відбувається автоматичне створення ієрархії (рис. 4.22):

- a. Якщо в робочій області не вибрано жодний об'єкт, то вставлені об'єкти не будуть зв'язані з об'єктами початкової онтології, і будуть розміщені в лівій верхній частині онтографу (1).
- b. Якщо вибраний один об'єкт початкової онтології, то кореневі об'єкти вставленого піддерева будуть зв'язані з ним (2).



- с. Якщо вибрано кілька об'єктів, то буде відображене попередження про створення великої кількості зв'язків. Якщо користувач натисне кнопку підтвердження, то кожен кореневий об'єкт вставленого піддерева буде зв'язаний з кожним з вибраних об'єктів (3).

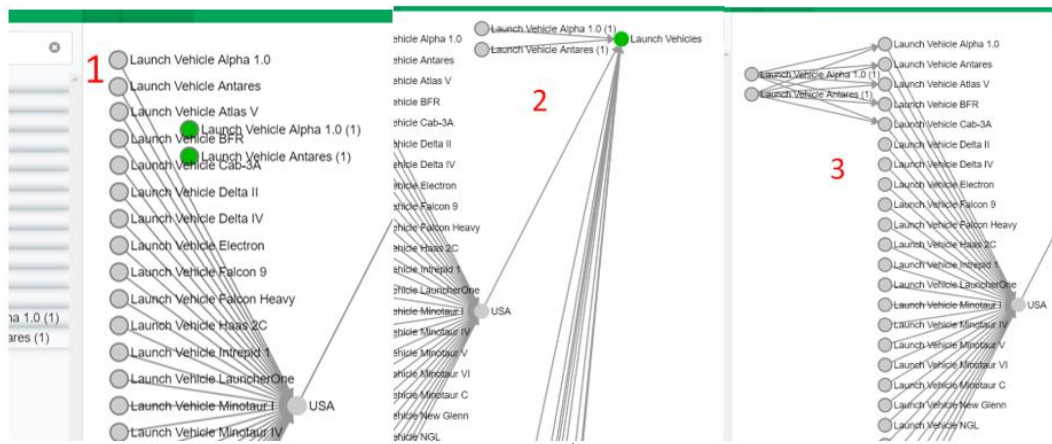


Рисунок 4.22 – Автоматичне створення ієрархії

Додатково компоненти редагування атрибутів підтримують копіювання атрибутів. Скопійовані таким чином атрибути будуть збережені в буфері обміну як один об'єкт з іменем «Data import». Цей об'єкт також може бути вставлений за описаною вище процедурою, але його потрібно буде перейменувати після вставки вручну.

#### 4.2. Онтологічний шаблон представлення для подальшої обробки результатів аналізу

Онтологічний шаблон представлення являє собою онтологію з одним або кількома вузлами, кожен з яких представляє конфігурацію. Конфігурація призначена для створення нових і конфігурації існуючих програмних компонентів.

Онтологія має розміщуватись на сервері – або як стандартна онтологія в публічній або приватній бібліотеці, або як файл в файлової системі серверу. Приклад XML-файлу онтологічного шаблону показано на рис. 4.23.

```

<Graph>
  <Nodes>
    <Node nodeName="API" nclass="" shape="circle" color="13421772" xPos="339" yPos="79">
      <data tclass="condition" type="text" link="">landing or internal</data>

      <data tclass="removeDefaultCSS" type="text" link="">1 or 0</data>
      <data tclass="stylesheet" type="href" link="EXT:grapheditor/Resources/Public/css/stub.css"></data>
      <data tclass="stylesheetLocalDir" type="href" link="EXT:grapheditor/Resources/Public/css/"></data>
      <data tclass="inlineStyle" link="" type="text">
        /*css*/
      </data>

      <data tclass="jsWhitelist" type="text" link="">/**regexp**</data>
      <data tclass="jsBlacklist" type="text" link="">/**regexp**</data>
      <data tclass="javascript" type="href" link="EXT:grapheditor_extensions/Resources/Public/js/stub.js"></data>
      <data tclass="javascriptLocalDir" type="href" link="EXT:grapheditor/Resources/Public/css/"></data>
      <data tclass="inlineScript" link="" type="text">
        jQuery(document).on('controlview.initialize', function (evt, that) {
          //js
        });
      </data>

      <data tclass="uncacheAssets" type="text">1 or 0</data>

      <data tclass="headerData" link="" type="text">
        headerData
      </data>

      <data tclass="templateWhitelist" type="text">/**regexp**</data>
      <data tclass="templateLocalDir" type="text" >EXT:grapheditor_extensions/Resources/Private/Handlebars/Stub</data>
      <data tclass="remoteTemplate" type="link" link="http://uri/of/template"></data>
      <data tclass="template:Stub" type="text"><![CDATA[
        html
      ]]></data>

      <data tclass="lang" type="group">
        <data tclass="lang_code" type="group">
          <data tclass="ua" type="text">ua</data>
          <data tclass="en" type="text">en</data>
        </data>
      </data>
    </Node>
  </Nodes>

```

Рисунок 4.23 - Приклад онтологічного шаблону

Онтологічний шаблон підключається до онтологічного сайту Адміністратором. Також Адміністратор встановлює онтологію за замовчуванням – онтологію, що відображається при первинному заході на онтологічний сайт (рис. 4.24).

**Plugin Options** [pi\_flexform]

Function    Конфігурація сайту    Публічна бібліотека

---

**Інформаційна онтологія по-замовчуванню**

https://e-devel.inhost.com.ua/?fname=nasu-property-descr

---

**Онтологічний шаблон представлення**

EXT:grapheditor\_extensions/Resources/Public/nasu-property/nasu-property.xml

---

Рисунок 4.24 - Інтерфейс Адміністратора

Використовуються наступні налаштування:

- 1) condition – умова застосування конфігурації:

- a. Не задана – конфігурація використовується завжди.
  - b. landing – конфігурація застосовується для «стартової» сторінки – тобто у випадку, якщо не вибрана онтологія (або вибрана онтологія за замовчанням).
  - c. Internal – конфігурація застосовується у випадку, якщо вибрана онтологія (що не являється онтологією за замовчуванням).
- 2) Налаштування для керування стилями (CSS):
- a. removeDefaultCSS – якщо має значення 1, то стандартні CSS-файли KIT не будуть включені взагалі. Це налаштування не впливає на файли вищого рівня (див. відповідний розділ даного документу).
  - b. stylesheet – дозволяє підключити файл CSS. Файл може розміщуватись на віддаленому сервері (використовується абсолютний URL) або локальний (відносний URL або TYPO3 file identifier).
  - c. stylesheetLocalDir – каталог, всі CSS-файли з якого мають бути підключеними (за виключенням вже підключених раніше – що дозволяє підключити файли в правильному порядку).
  - d. inlineStyle – дозволяє вказати CSS-код прямо в онтології. Даний код буде розміщено після всіх підключених CSS-файлів, тобто він матиме найвищий пріоритет.
- 3) Налаштування для керування виконуваними файлами (JavaScript):
- a. jsWhitelist – регулярний вираз в форматі PCRE, що дозволяє вказати ті СТАНДАРТНІ JavaScript – файли, що будуть підключені. Дозволяє значно зменшити об'єм даних, що передається з сервера, але таким чином дуже легко зробити модуль непрацездатним. За додатковими деталями зверніться до відповідного розділу даного документу.
  - b. jsBlacklist – регулярний вираз в форматі PCRE, що дозволяє вказати ті СТАНДАРТНІ JavaScript – файли, що НЕ будуть підключені. За додатковими деталями зверніться до відповідного розділу даного документу.

- c. `javascript` – дозволяє підключити файл JavaScript. Файл може розміщуватись на віддаленому сервері (використовується абсолютний URL) або локальний (відносний URL або TYPO3 file identifier).
  - d. `javascriptLocalDir` – каталог, всі JavaScript-файли з якого мають бути підключеними (за виключенням вже підключених раніше – що дозволяє підключити файли в правильному порядку).
  - e. `inlineScript` – дозволяє вказати JavaScript -код прямо в онтології. Даний код буде розміщено в заголовку після jQuery, тому для коректної роботи в ньому слід використовувати події (див. відповідний розділ даного документу).
- 4) `uncacheAssets` – у випадку, якщо CSS і JS файли розміщуються локально, КІТ знає час останньої їх зміни і генерує посилання таким чином, щоб браузер не використовував застарілу версію файлу з кеша. Для віддалених файлів це неможливо, тому використовується час останньої зміни онтології. Встановлення даного значення в 1 відключає кешування віддалених файлів, і таким чином вирішує потенційно зв'язані з цим проблеми. Негативною стороною є зменшення швидкодії через необхідність кожен раз завантажувати файли з віддаленого сервера.
- 5) `headerData` – дозволяє додати довільний HTML-код в заголовок (`<header>`) сторінки. Може використовуватись для додавання спеціалізованих елементів, таких, як OpenGraph-метатеги.
- 6) Налаштування для керування HTML-шаблонами:
- a. `templateWhitelist` – регулярний вираз в форматі PCRE, що дозволяє вказати ті коди шаблонів, які будуть підключені. Дозволяє значно зменшити об'єм переданих з серверу даних.
  - b. `templateLocalDir` – каталог, з якого будуть завантажені HTML-файли шаблонів (код шаблону – назва файлу).
  - c. `remoteTemplate` – звантажує HTML-файл шаблону з віддаленого серверу. Ресурсозатратна функція, використання якої слід мінімізувати.

- d. `template:TemplateCode` – обробляє текстовий вміст елемента даних як шаблон з кодом `TemplateCode`. Текстовий вміст є HTML, і тому повинен бути розміщений всередині `![CDATA[...]]>`.
- 7) `lang` – дозволяє задати додаткові мовні змінні. Дане налаштування повинно мати тип `group`.
  - a. Дочірніми до елементами до групи `lang` є групи `<data tclass="lang_code" type="group">`, де `lang_code` – код мовної змінної.

Дочірніми до елементів `lang_code` є текстові елементи, `tclass` яких – код мови (`en` або `ua`), а текстовий вміст – значення мовної змінної у відповідній мові.

### 4.3. Модуль класифікації та обробки інформації

#### 4.3.1. Загальна структура модулю класифікації та обробки інформації

Даний програмний модуль, є розширенням модулю «Рекурсивний редуктор» із стандартної бібліотеки модулів КІТ, доступ до якого надається адміністратором.

Перед початком роботи з Редуктором необхідно перевірити наявність встановлених на ПК наступних компонентів:

- 1) Середовище `java virtual machine`;
- 2) Веб-браузер.

Всі виконувані файли Редуктора містяться в робочій папці, яка розповсюджується єдиним пакетом. Загальна структура робочої папки редуктора представлена на рис. 4.25.

Папка містить наступні елементи:

- 1) Каталог з редактором правил (`editor`);
- 2) Каталог з вхідними даними (`sources`);

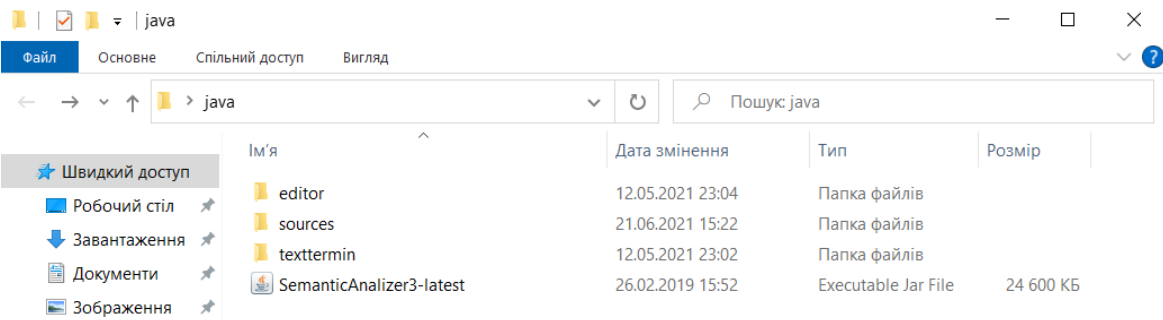


Рисунок 4.25 - Структура робочої папки редуктора

3) Каталог з лексичним аналізатором [109] (Програмний модуль «Конспект (TextTermin)») виконує лінгвістичний аналіз тексту до рівня поверхневого синтактико-семантичного аналізу;

4) Виконуваний файл редуктора.

Каталог з вхідними даними створюється окремо для кожної конкретної задачі.

Структура каталогу включаю наступні елементи (рис. 4.26) :

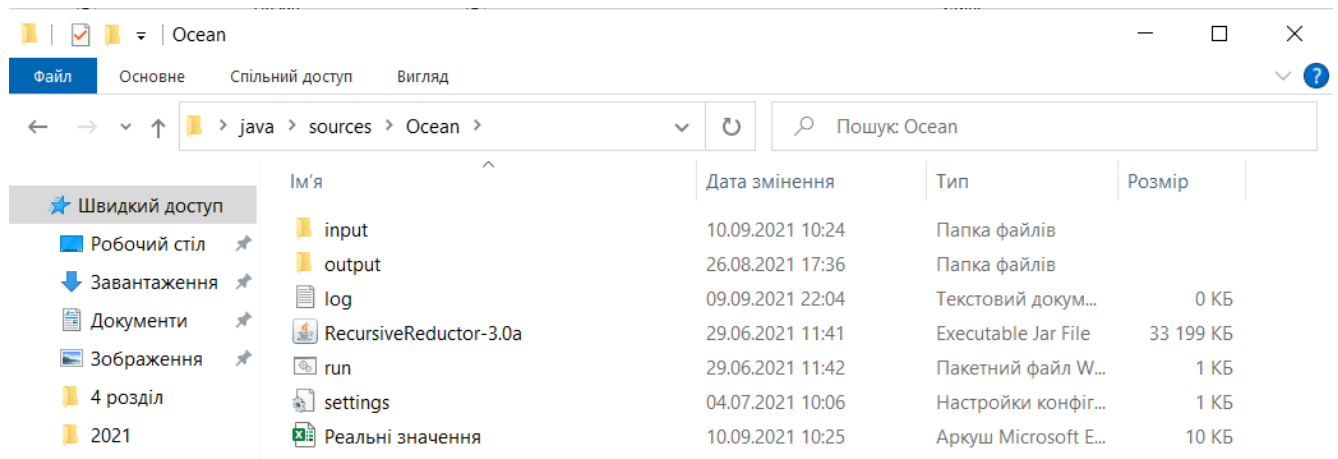


Рисунок 4.26 - Структура каталогу з вхідними даними

- 1) Вхідні файли (у форматі \*.docx) (input);
- 2) Вихідні файли (output);
- 3) (log);
- 4) Виконуваний модуль (RecursiveReductor-3.0a);
- 5) bat-файл виконання обробки (run);
- 6) Файл конфігурації (settings);
- 7) Файли з додатковими даними (Реальні значення);

Файл конфігурації – це стандартний INI-файл, який має секції та їхні значення розділені знаком «дорівнює». Для створення файлу конфігурації доступні такі секції:

writerModule – основний модуль обробки;

allowEmptyInput – дозволяє модулю самому вибирати вхідні файли;

rootNodeName – коренева вершина онтології;

configPrimary – керуюча онтологія;

workDir – робочий каталог системи;

ontoconfig – дозволяє прописувати значення на які посилається онтологія;

defaultAPI – інформація для з'єднання з Поліедром через Web;

defaultCache – кешування результатів обробки файлів перед передачею їх обробки;

inheritance – батьківські налаштування;

realValues – додаткові дані для обробки;

debug – вмикає вивід додаткової інформації про процес обробки

input – вхідні файли обробки;

output – вихідні файли.

Вхідні і вихідні файли можуть бути чотирьох видів:

api – файли зчитуються або завантажуються на сервер КІТ, заданий секцією defaultAPI. Значення повинно бути назвою або ідентифікатором онтології

local – файли зчитуються або зберігаються в рамках локальної файлової системи. Значення повинно містити шлях до файлу.

http – файли зчитуються з ресурсу в мережі Інтернет. Значення повинно містити посилання (тільки для секції input).

Sftp – файли зчитуються або завантажуються на віддалений сервер з допомогою протоколу SSH File Transfer Protocol.

Приклад файлу конфігурації для обробки океанографічних законодавчих актів наведено на рис. 4.27.

```

settings: Блокнот
Файл Редагування Формат Вигляд Довідка
[[general]
writerModule = SvitgaiOntologyBased
rootNode= Нормативний документ

debug = 1

ontoConfig = sharedgraph=60d08aec7c58d
realValues = Реальні значення.xlsx
mode=classification

[remote]
url=https://e-devel.inhost.com.ua/editor/

[input]
local = input/

[output]
local = output/Закон.csv

```

Рисунок 4.27 - Файл конфігурації обробки законодавчих актів

## 4.4. Трансдисциплінарна інтеграція інформації

### 4.4.1. Робота з індексними зонами в середовищі КІТ «Поліедр»

Трансдисциплінарна інтеграція інформації здійснюється в рамках індексних зон, кожна з яких представлена певним сервісом типу «Індексатор». Індексні зони можуть задаватись або в налаштуваннях додатку, або в налаштуваннях користувача. Вікно керування індексаторами відкривається відповідною кнопкою в головному меню (рис. 4.28). Користувач повинен мати право на керування хоча б одним індексатором, інакше кнопка буде неактивною.

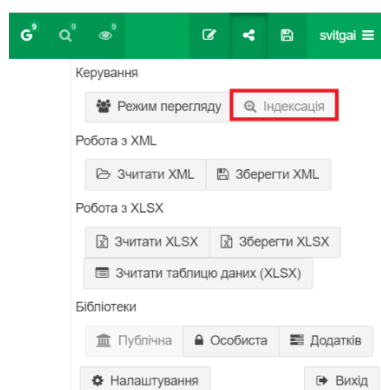


Рисунок 4.28 - Кнопка індексації в головному меню Редактора онтологій

Вікно керування використовується для внесення інформаційних ресурсів в індекс, і містить наступні керуючі елементи (рис. 4.29):



1) Випадаючий список вибору індексатора – при зміні індексатора вікно оновлюється, щоб показати налаштування вибраного індексатора.

2) Глибина кравлінгу – визначає, як глибоко буде індексатор спускатись вниз по дереву гіперпосилань. 0 означає, що індексується тільки поточна сторінка, 1 – поточна сторінка і сторінки, на які є посилання на поточній, і т.д.

3) Теги – список стандартних (заданих адміністратором) тегів для даного індексатора (3). При виборі одного або кількох елементів зі списку (необхідно утримувати Ctrl) вибрані теги будуть присвоєні всім інформаційним ресурсами, що будуть внесені в індекс в результаті операції (в залежності від значення параметру «Глибина кравлінгу»). Теги можуть відображатись як вкладені (4), однак ця функція надає виключно візуальну структуру і не має ніякого впливу на роботу пошуку чи індексації.

4) Додаткові теги (5) – дозволяє додавати теги, відсутні в заданому адміністратором списку. Використання таких тегів при пошуку можливе, якщо додати в пошуковий запит текст виду `#*назва тегу*`, але в цілому використання додаткових тегів є небажаним.

5) Посилання на веб-ресурси – власне посилання на інформаційні ресурси, що будуть проіндексовані. Можна виділити основні типи ресурсів:

- a. Сторонні сайти (6) – можна вказати будь-яке валідне посилання. Однак слід розуміти, що сайти можуть мати певні обмеження і квоти, що не даватимуть проіндексувати їх достатньо ефективно. Необхідно визначати глибину кравлінгу згідно з розумінням структури ресурсу, так, якщо це звичайний файл – слід вказати 0.
- b. Окремі онтології (7) – повинні бути в рамках тої ж інсталяції, де буде розміщений інтерфейс пошуку. Посилання на онтологію можна скопіювати в адресному рядку браузера, а при індексації слід виставляти глибину кравлінгу 2.
- c. Онтологічні сайти (8) – необхідно вставити посилання на головну сторінку сайту, і додати до неї параметр `list=1`, щоб відобразився

список доступних в рамках сайту онтологій. Глибина кравлінгу повинна дорівнювати 3.

6) Власне кнопка індексації (9) запускає індексацію з заданими параметрами.

7) Кнопка перевірки стану індексації (10) дозволяє визначити, чи відбувається в даний момент індексація. Після натиску буде відображене повідомлення з результатом перевірки: процес індексації триває, індексація завершена, індексатор не підтримує функцію перевірки, індексатор недоступний.

8) Кнопка очистки індексу (11) дозволяє повністю очистити індекс. Це буває доцільним у випадку, коли в індексі є нерелевантні або застарілі інформаційні ресурси, які необхідно виключити з видачі.

Керування індексаторами

Індексатор  
1 Індексатор - ТЕСТ

Глибина кравлінгу  
2 2 (значення для індексації окремої онтології)

Теги  
3 #tag1  
4 #tag2  
-#tag2-1  
-#tag2-2

Додаткові теги  
5 Розділені пробілом значення без символу #

Посилання на веб-ресурси  
Один рядок – одне посилання. При індексації онтосайтів необхідно додати параметр: <https://museum.stemua.science/?list=1>  
6 <https://museum.stemua.science/?fname=klasifikacija-muzejiv-za-prinalezhnistju-do-zakladiv-osviti>  
7 <https://example.com>  
8 <https://museum.stemua.science/?fname=klasifikacija-muzejiv-za-osobistostjami>

9 Індексувати 10 Перевірити 11 Очистити індекс

Рисунок 4.29 – Вікно керування індексаторами

#### 4.4.2. Пошук та представлення інформації в середовищі КІТ «Поліедр»

Пошук інформації здійснюється з допомогою блоку пошуку, що розміщується в верхній панелі додатку а також в заголовку вікна детального перегляду атрибутів (рис. 4.30).

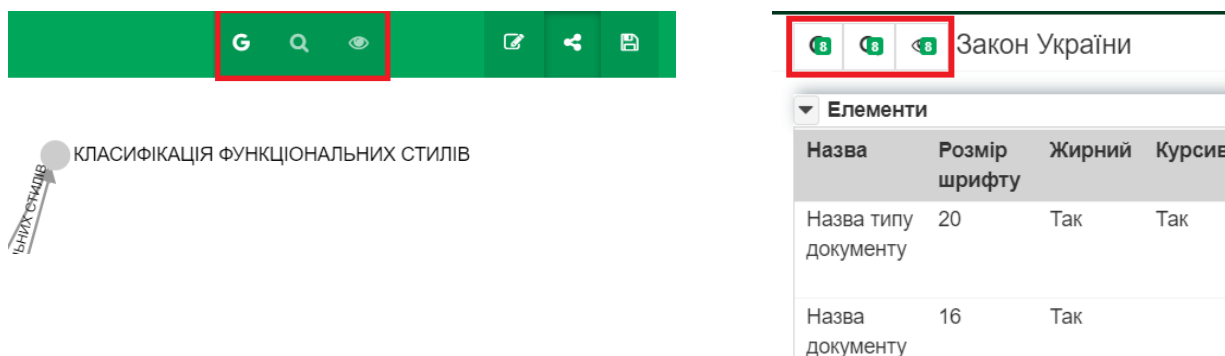


Рисунок 4.30 - Блок пошуку інформації

Кожен з елементів блоку запускає один з режимів пошуку. При запуску пошуку в якості пошукового запиту використовується вибраний в даний момент текст. Механізм вибору має особливості:

- Більшість тексту, що відображається в рамках додатку, доступна для виділення мишкою і автоматично перетворюється в пошуковий запит за замовчуванням;
- Якщо ввімкнено режим перегляду онтографу, то виділення об'єкту в робочій області використає його ім'я в якості пошукового запиту за замовчуванням. Якщо вибрано кілька об'єктів, то їх імена об'єднуються через пробіл;
- Довжина поточного вибраного тексту (в символах) показується у вигляді відміток на кнопках блоку пошуку. Якщо довжина перевищує максимальний розмір пошукового запиту (за замовчуванням – 200), то запит буде обрізаний, а відмітки будуть виділені жовтим.

Першою в блоці пошуку відображається кнопка **зовнішнього пошуку** – Натиск на неї відкриє сторінку Google з вибраним поточним пошуковим запитом. Якщо запит пустий, кнопка не спрацює і буде показане попередження.

Зовнішній пошук застосовується достатньо рідко. Як правило, пошук здійснюється всередині індексних зон, для чого використовуються компоненти внутрішнього і онтологічного пошуку, а в спеціалізованих випадках (для трансдисциплінарної консолідації) – додаток «Пошукова призма».

**Заголовок компоненту пошуку** – це набір керуючих елементів, що використовуються компонентами внутрішнього і онтологічного пошуку (рис. 4.31) і містить:

1) Власне поле для вводу пошукового запиту з кнопками виконання пошуку і очистки пошукового запиту.

2) Елемент вибору індексної зони – при його зміні і непустому полі пошукового запиту автоматично запускається індексація.

3) Елемент фільтрації за датою індексації – для індексаторів, що підтримують дану функцію.

4) Елемент вибору мови пошуку – для індексаторів, що підтримують багатомовність.

5) Кнопку відкриття додаткових параметрів – вона розміщується зліва від елементу вибору індексної зони і при натисканні показує/приховує елементи фільтрації за датою і вибору мови пошуку.

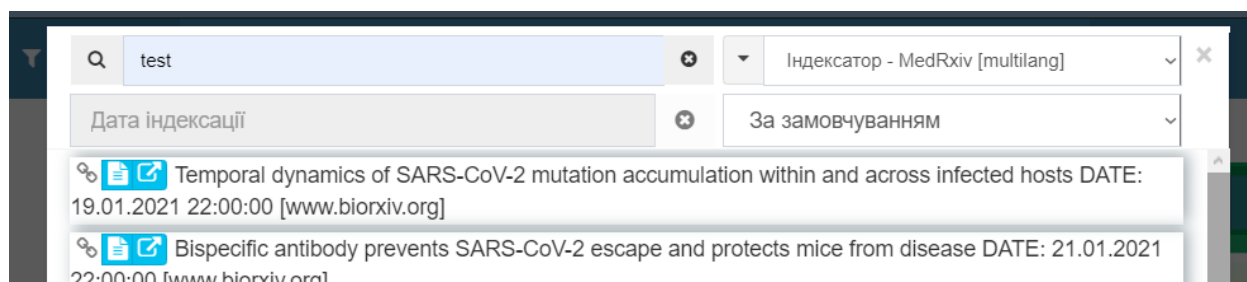


Рисунок 4.31 - Заголовок компоненту пошуку

Вікно компоненту внутрішнього пошуку, що викликається при натисканні відповідної кнопки, має наступний вигляд (рис. 4.32). Особливості роботи компоненту:

- Пошукові результати відображаються у вигляді лінійного списку, розділеного на сторінки (50 результатів на сторінку). Перемикання сторінок здійснюється елементами внизу списку;
- Кожен елемент містить:
  - Індикатор типу результату – результат може бути файлом, посиланням або об'єктом онтології. Індикатор показує відповідну іконку типу а при наведенні на нього відображається назва типу;

- Кнопку «Деталі» – відкриває вікно детального перегляду пошукового результату;
- Кнопку «Перехід» – переходить за посиланням, за яким розміщено отриманий в результаті пошуку об’єкт;
- Назву пошукового результату – назву файлу, заголовок сторінки чи ім’я об’єкту.

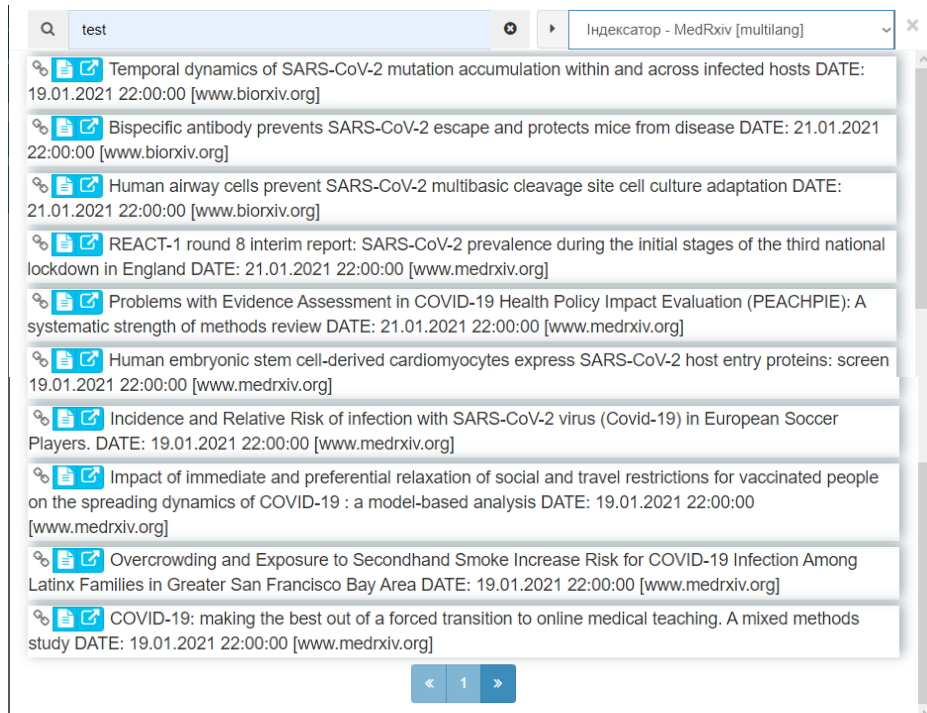


Рисунок 4.32 - Вікно компоненту внутрішнього пошуку

Компонент онтологічного пошуку призначений для відображення пошукових результатів у вигляді онтографу (рис. 4.33).

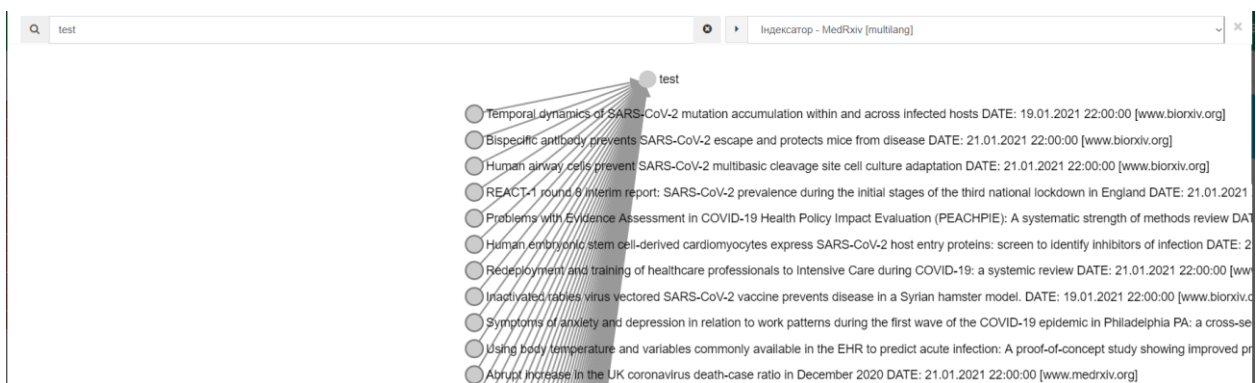


Рисунок 4.33 - Компонент онтологічного пошуку

Основною перевагою онтологічного пошуку є те, що з ним працюють деякі функції представлення онтографу:

- Подвійний клік на об'єкт відкриває вікно детального перегляду пошукового результату, як і у вікні внутрішнього пошуку;
- Одинарний клік на об'єкті вибирає його, натискання Shift дозволяє вибрати множину об'єктів;
- Натиск Ctrl+A виділяє всі видимі об'єкти;
- Натиск Ctrl+C виконує копіювання виділених об'єктів в буфер обміну.

Функція копіювання дозволяє створювати онтології на основі пошукових запитів, детальніше ця функція описана в Інструкції експерта ППГ.

Обидва компоненти пошуку викликають вікно детального перегляду результату пошуку, що призначене для відображення користувачу способу вживання слів з пошукового запиту в тексті пошукового результату. При цьому логіка роботи вікна різна для структурованих і неструктурованих результатів.

Для структурованих результатів (результатів типу «об'єкт онтології») використовується представлення, аналогічне до вікна детального перегляду атрибутів (рис. 4.34, зліва). Всі атрибути об'єкту відображаються у вікні, а входження слів з пошукового виділяються жовтим. При цьому якщо входження міститься всередині складної структури даних (групи, таблиці), то відповідне піддерево буде розкрито.

Для неструктурованих результатів (файл і посилання) використовується спрощене представлення (рис. 4.34, справа), де контексти слів з пошукового запиту відображаються як текстові елементи.

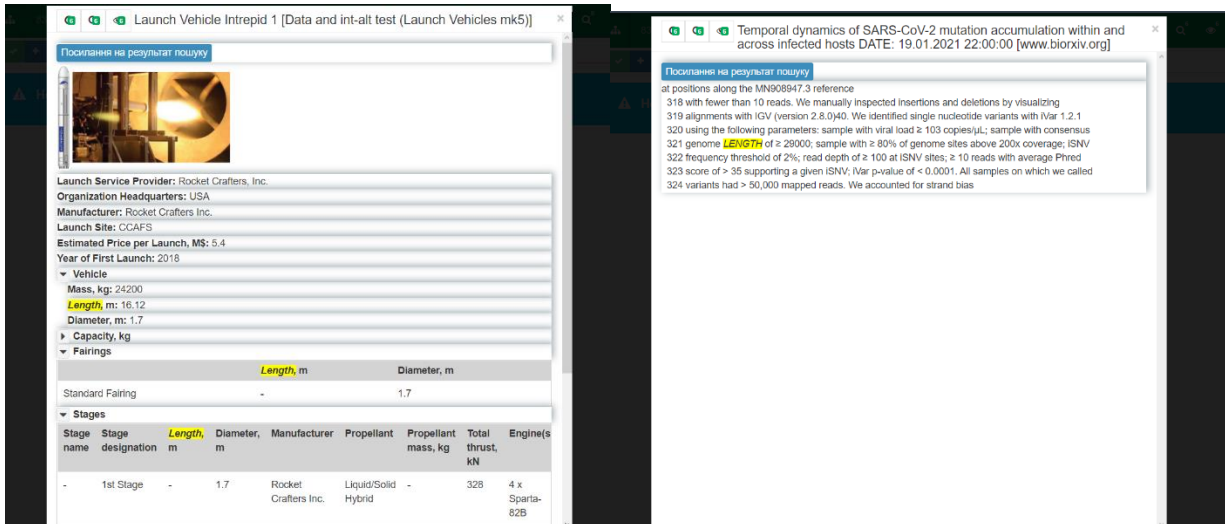


Рисунок 4.34 - Вікно детального перегляду для структурованих і неструктурованих результатів

Пошукова призма являє собою окремий додаток, призначений для трансдисциплінарної консолідації інформації, що міститься у великих масивах мережових інформаційних ресурсів (рис. 4.35).



Рисунок 4.35 - Загальний вигляд додатку «Пошукова призма»

Пошукова призма є варіантом онтологічної призми, в якому використовуються тільки два рівня онтології (об'єкт-призма і об'єкт-грань). При цьому грані являються пошуковими запитамі, а їх вміст (третій рівень) представлений пошуковими результатами.

Особливості роботи додатку «Пошукова призма»:

- Верхня панель виділена темно-синім кольором;
- Вимкнено атрибутивний фільтр, а з режимів перегляду є тільки режим онтографу і власне пошукової призми.

Режим перегляду пошукової призми відрізняється від режиму звичайної онтологічної призми вмістом її граней, а також наявністю спеціалізованого віджета, що містить стандартний перемикач індексних зон, і перемикач «Видаляти пусті грані». Перемикач дозволяє ховати грані, для яких не знайдено жодного пошукового результату. За замовчуванням перемикач вимкнено, оскільки пусті грані можуть представляти вкладені пошукові призми, в яких результати можуть бути присутні. Такі грані відображаються так, як показано на рис. 4.36.

Додатково в рамках додатку працюють всі стандартні функції пошуку, наприклад, якщо грань не представляє вкладену призму, то користувачу зразу буде запропоновано знайти інформацію за даним пошуковим запитом у внутрішньому пошуку.

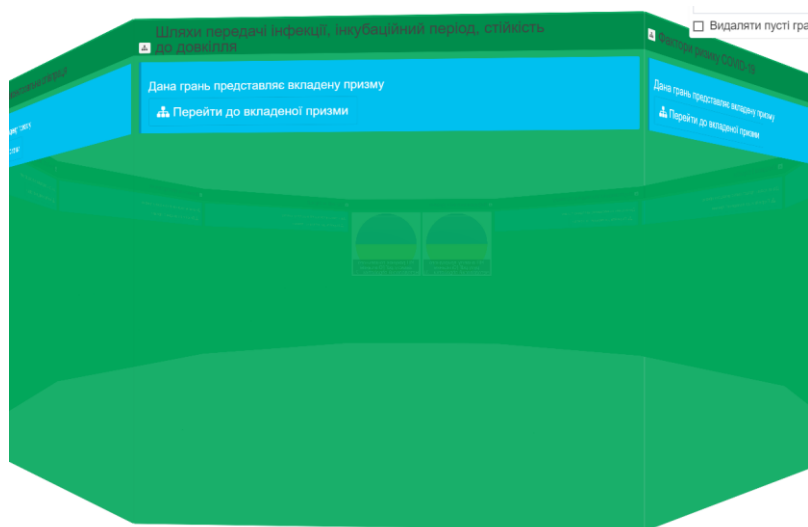


Рисунок 4.36 - Пуста грань, що представляє вкладену призму

#### 4.5. Приклад трансдисциплінарного представлення великого масиву неструктурованих текстів

В якості великого масиву неструктурованих текстів обрано: інформаційну базу знань наукової продукції (ІБЗ НП), навчальні програми, а також навчально-дослідницькі проєкти НЦ «МАНУ». Метою трансдисциплінарного представлення цих інформаційних ресурсів є:



– Консолідація інформаційних ресурсів, що відображають науково-технічну продукцію (НТП) з навчальними програмами для визначення якості і повноти наявної в ІБЗ НП системи знань;

– Консолідація інформаційних ресурсів, що відображають науково-технічну продукцію (НТП) з навчально-дослідницькими проектами НЦ «МАНУ» для визначення якості цих проектів.

В обох випадках консолідація здійснюється між ресурсами в рамках ІБЗ НП, що означає необхідність внесення в неї програм і проектів відповідно. Оскільки обидва типи документів є слабо структурованими, витриманими в різних стилях, до них спочатку застосовуються технологічні засоби структуризації.

Приклад навчальної програми, що обробляється, показано на рис. 4.37.

**Компетентнісний потенціал навчального предмета**

<p>1. Спілкування державною (і рідною у разі відмінності) мовами</p>	<p><b>Уміння:</b> усно й письмово тлумачити біологічні поняття, факти, явища, закони, теорії; описувати (усно чи письмово) експеримент, послуговуючись багатим арсеналом мовних засобів — термінами, поняттями тощо; обговорювати проблеми біологічного змісту.</p> <p><b>Ставлення:</b> усвідомлення значущості здобутків біологічної науки, зокрема пошанування досягнень українських учених; прагнення до розвитку української біологічної термінологічної лексики.</p> <p><b>Навчальні ресурси:</b> навчальні, науково-популярні, художні тексти про природу, дослідницькі проекти в галузі біології, усні / письмові презентації їх результатів</p>
--	--

Рисунок 4.37 – Приклад фрагменту навчальної програми

Така програма містить велику кількість інформації, яка в цілому зводиться до переліку ключових слів. Такі ключові слова необхідно виділити з тексту програми, класифікувати і налаштувати процеси автоматизованої оцінки наявних в ІБЗ НП інформаційних ресурсів на її основі.

Оскільки в рамках поставленої задачі структура самої програми не має значення, це дозволяє обробляти їх за спрощеною схемою – з допомогою

лексикографічних модулів. Відповідний дескриптор структуризації показано на рис. 4.38.

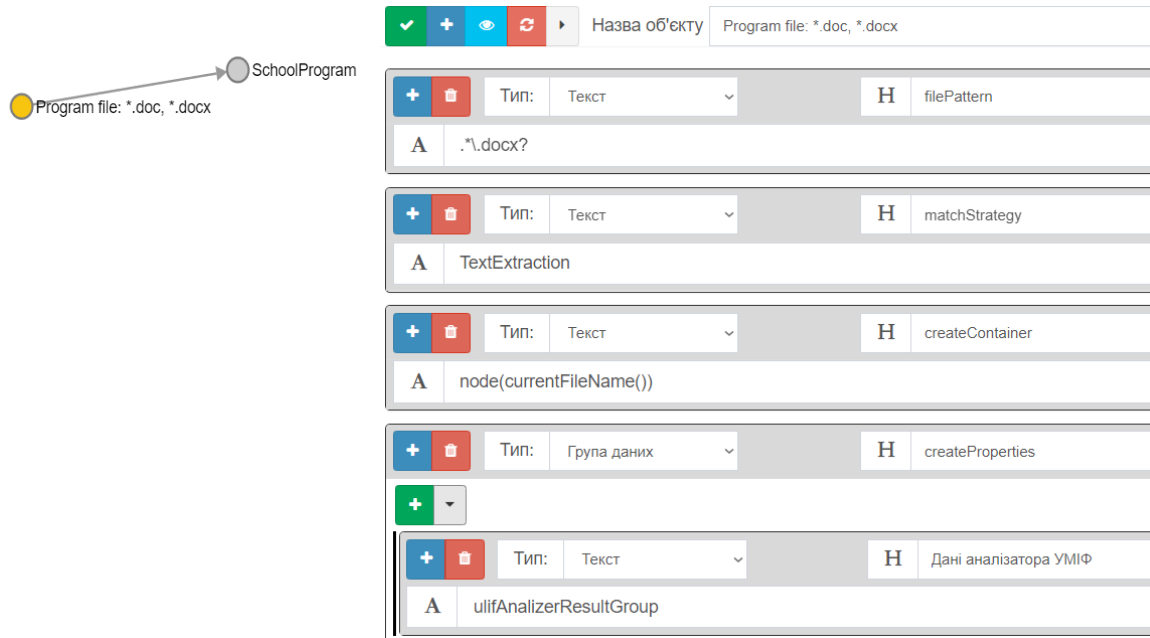


Рисунок 4.38– Онтологічний дескриптор структуризації навчальної програми

Даний варіант дескриптора дозволяє робити одну інформаційну онтологію на основі масиву програм. В результаті утворюється онтологія (рис. 4.39), що містить самі програми в якості об'єктів, а отримані з них терміни – в якості атрибутів. Як вже було показано вище, експерт може використовувати такі онтології разом з наявними механізмами повнотекстового пошуку і встановлювати ступінь відповідності наявних матеріалів програми.

За результатами такого аналізу доповнюються або модифікуються наявні в ІБЗ НП інформаційні ресурси.

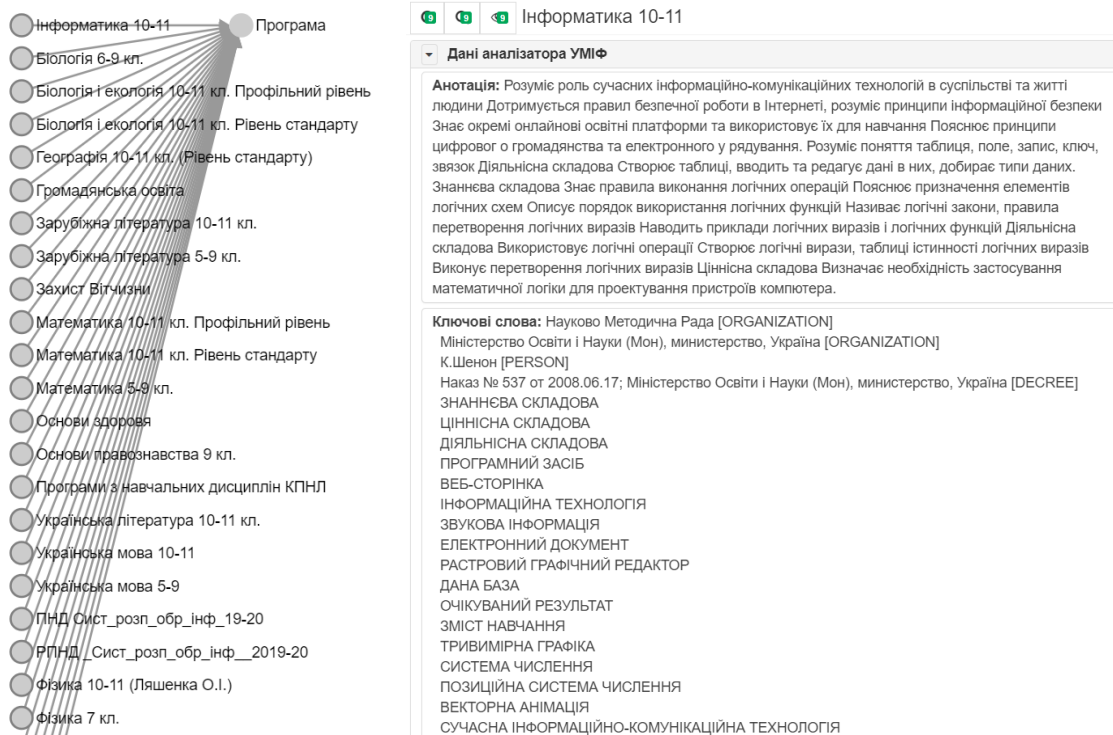


Рисунок 4.39 – Результат структуризації масиву навчальних програм

Для навчально-дослідницьких проєктів НЦ «МАНУ» використовується дещо інший підхід, що враховує структуру документу. Крім власне ключових слів, що також виділяються з таких проєктів, аналізується також структура самого документу та його складові частини.

Приклад обробки навчально-дослідницького проєкту показано на рис. 4.40.

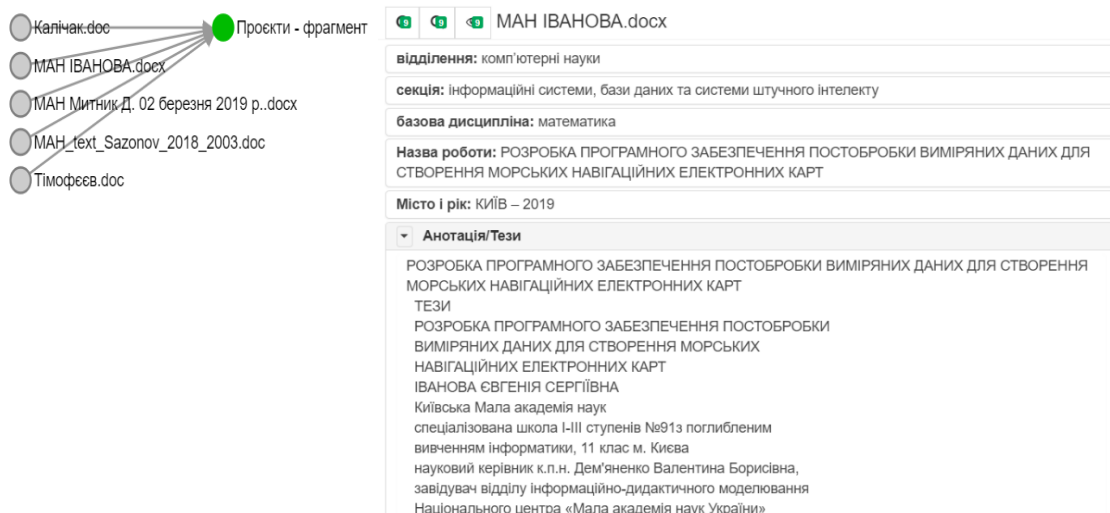


Рисунок 4.40 – Приклад обробки навчально-дослідницького проєкту

При створенні онтології до атрибутів об'єкту додається власне вихідний текст учнівського проєкту, розбитий на атрибути типу «Група» у відповідності до вкладеності розділів. Таким чином при аналізі роботи експерт може переглядати не тільки ключові слова, але і контекст, в якому ці ключові слова вживаються в роботі. Такий підхід забезпечує більш повне відображення роботи.

Вказані вище термінологічні онтології не є кінцевим результатом. Вони є першим кроком до ефективного представлення знань про НТП. Другим кроком є формування онтології пошукового запиту, що може здійснюватися рядом способів.

Найпоширеніший зі способів – це обробка термінологічної онтології експертно (рис. 4.41). Експерт відбирає з наявного списку ті терміни, які він вважає релевантним своїй задачі, і вносить їх в електронну таблицю.

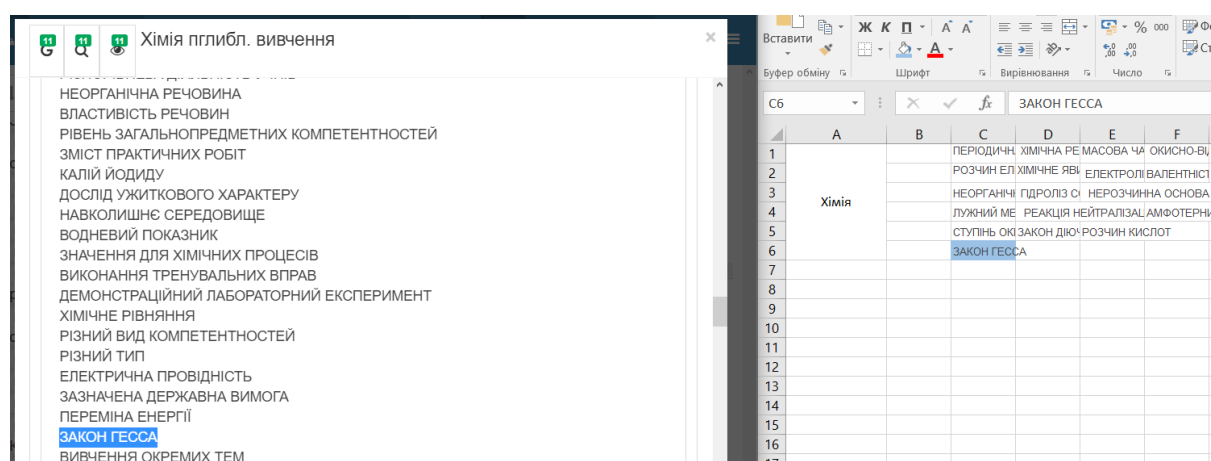


Рисунок 4.41 – Формування онтології пошукового запиту

Сформована онтологія (рис. 4.42) відображає бачення експерта на структуру його задачі. В залежності від цієї структури, вона може мати певну ієрархію або бути однорівневою (як показано на рисунку).

Така онтологія використовується для консолідованого представлення інформації про НТП в форматі онтологічної пошукової призми (рис. 4.43). Кожна грань такої призми представляє певний елемент ПГ в рамках задачі експерта, а елементи на грані призми представляють інформаційні ресурси, релевантні даному елементу.

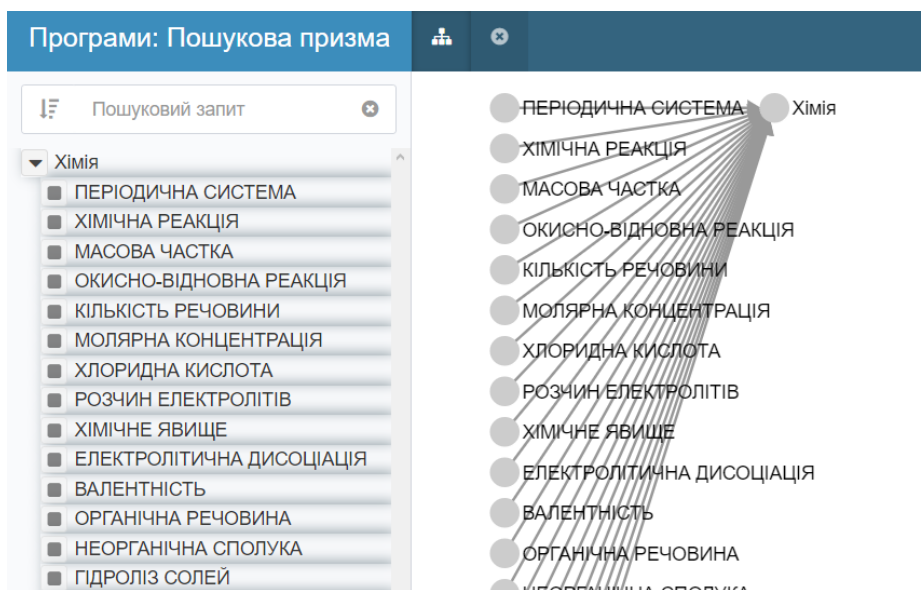


Рисунок 4.42 – Онтологія пошукового запиту

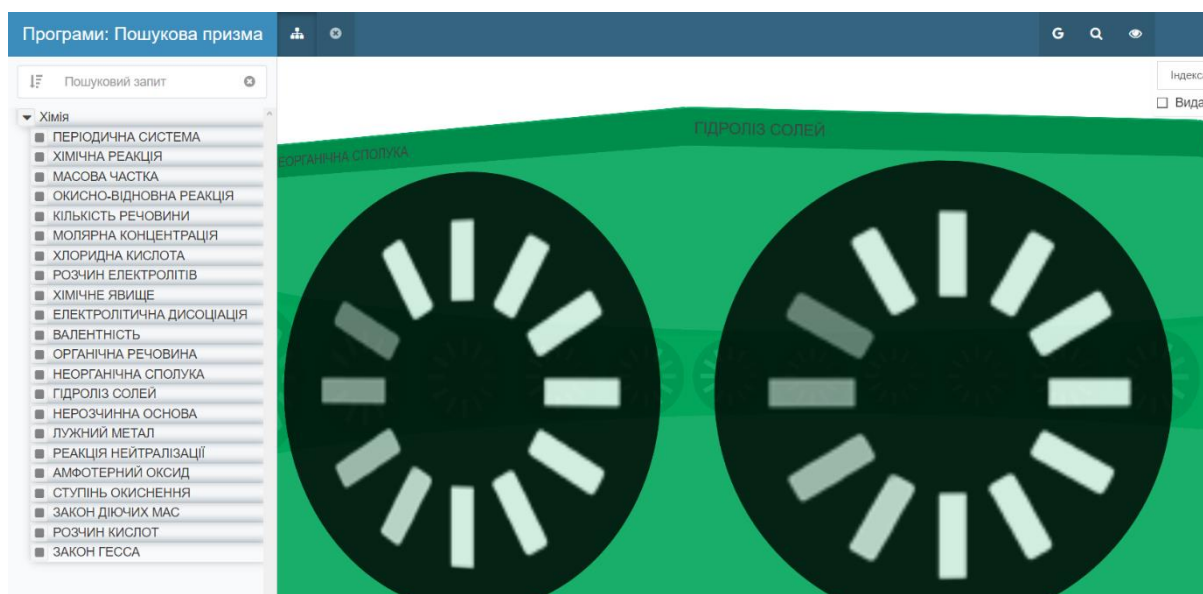


Рисунок 4.43 – Онтологічна пошукова призма НТП

#### 4.6. Приклад класифікації та вибору контенту

Існує багато репозиторіїв, які містять множини наративів, зміст яких відображає наукову продукцію. Яскравим прикладом є наукові сайти – medrxiv.org та biorxiv.org, які містять більш ніж 400000 наукових, науково-технічних та науково-методичних результатів щодо діагностування та лікування різних хвороб, включаючи COVID-19.

На основі результатів досліджень, що описано вище, створено «Кластер знань про коронавірусну інфекцію», який є мережевим, трансдисциплінарним

інформаційним ресурсом, що всеосяжно відображає існуючі на сьогодні дані щодо COVID-19.

Сервіс доступний за посиланням: <https://covid19tdm.ulif.org.ua/> і має загальний вигляд, як зображено на рис. 4.44.



Рисунок 4.44 – Загальний вигляд Кластеру знань про коронавірусну інфекцію

Кластер містить онтології пошукових запитів, що відповідають на питання, сформульовані експертами Білого дому США і Національної медичної академії післядипломної освіти імені П.Л. Шупика Міністерства охорони здоров'я України (рис. 4.45, рис. 4.46).

Кожний елемент кластеру представлений пошуковою призмою, яка забезпечує інтегрований доступ до агрегованих просторово розподілених цифрових активів про COVID-19 на основі вбудованих мовноінваріантних інструментів семантичного аналізу та автоматичної і динамічної структуризації мережевих інформаційних ресурсів. Для переходу до будь-якої пошукової призми необхідно навести на неї мишкою та натиснути лівою кнопкою.



Рисунок 4.45 – Онтологія пошукових запитів, що відповідають на питання, сформульовані експертами Білого дому США

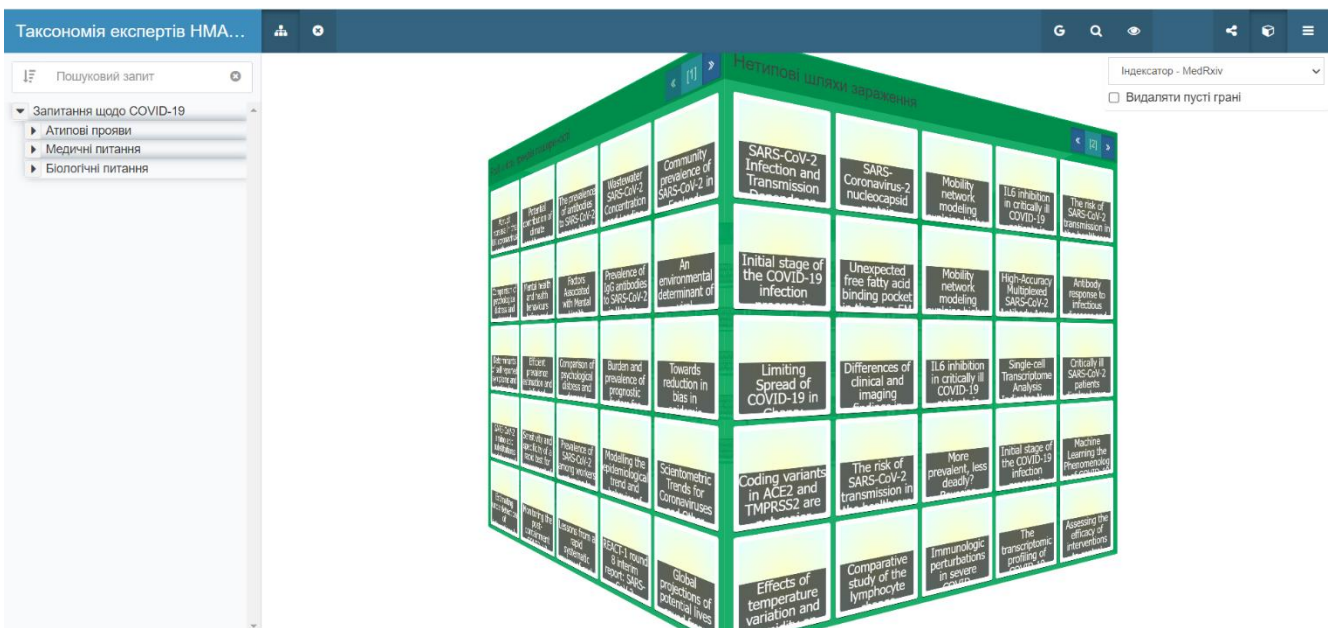


Рисунок 4.46 – Онтологія пошукових запитів Національної медичної академії післядипломної освіти імені П.Л. Шупика

Дані онтології мають складну структуру вкладених призм. На рис. 4.47 зображено піддерево онтології пошукових запитів Національної медичної академії, яке відповідає вершині «Медичні питання» .

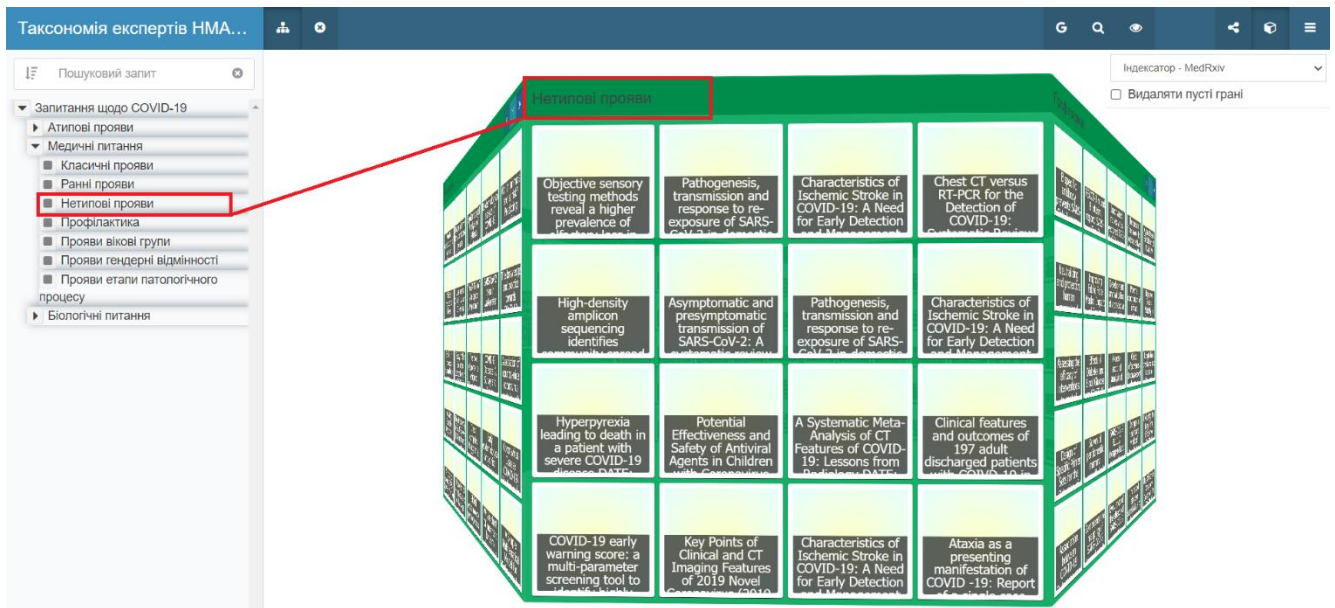


Рисунок 4.47 – Вкладена призма

Кожна грань даних пошукових призм відображає результати автоматичного вибору документів, що містять інформацію, відповідну назві даної грані (тобто заданої тематики). При виборі конкретного документу на грані, відкривається його опис (рис. 4.48) з можливістю переходу, за необхідності, безпосередньо до тексту документу (рис. 4.49). Перехід до тексту здійснюється шляхом натискання кнопки «Посилання на результат пошуку».

Air pollution, SARS-CoV-2 transmission, and COVID-19 outcomes: A state-of-the-science review of a rapidly evolving research area DATE: 19.08.2020 21:00:00 [www.medrxiv.org]

index values, and 3 studies used the number of daily limit exceedances.

The short-term time series studies used daily air pollution data collected over relatively shorter periods of time, ranging from 17 days to 80 days. These studies used daily air pollution data during the outbreak with a short lag time to account for **INCUBATION PERIOD** and/or time from exposure to death. However, Bashir et al. (2020) did not discuss the use of a lag period in their analyses, which should be taken into consideration when interpreting the results from that study. Eight of the short-term time-series studies used air pollutant concentrations or air quality index values, and

1384, 10.1021/acs.est.9b03358

Diamond M, Peniston Feliciano HL, Sanghavi D, Mahapatra S. 2020. Acute Respiratory Distress Syndrome (ARDS). Available: <https://www.ncbi.nlm.nih.gov/books/NBK436002/> [accessed August 3 2020].

Domingo JL, Rovira J. 2020. Effects of air pollutants on the **TRANSMISSION** and severity of respiratory viral **INFECTIONS**. Environmental Research 187:109650, <https://doi.org/10.1016/j.envres.2020.109650>

Dominici F, Zeger SL, Samet JM. 2000. A measurement error model for time-series studies of air pollution and mortality. Biostatistics 1:157-175, 10.1093/biostatistics/1.2.157

Dominici F, Zigler C. 2017. Best Practices

[Посилання на результат пошуку](#)

Рисунок 4.48 – Опис одного з документів



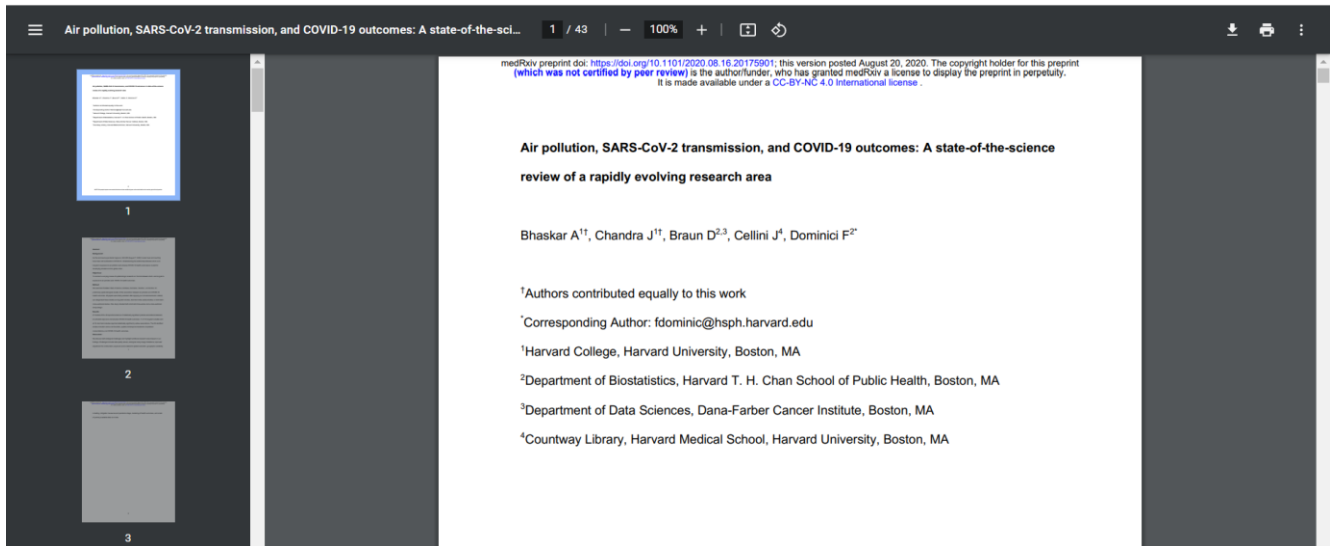


Рисунок 4.49 – Текст знайденого документу

Експерт може сформулювати власну онтологію пошукового запиту. Для цього авторизований користувач використовує спеціалізовані таблиці. З їх допомогою формуються об'єкти таксономії і зв'язки між ними. Таким чином створюється власна термінологічна таксономія (рис. 4.50) і реалізується класифікація документів.

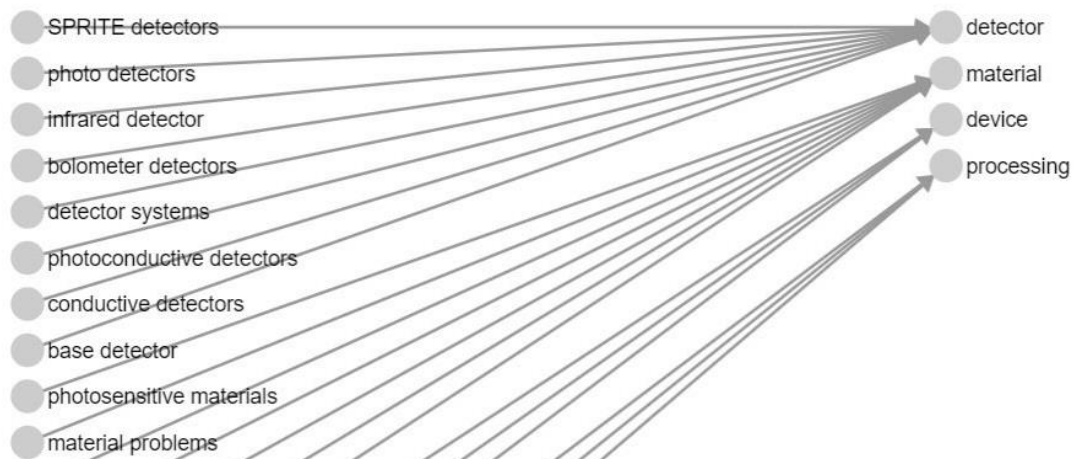


Рисунок 4.50 – Фрагмент експертної онтології пошукових запитів

Відповідно до кожного об'єкту експертної онтології пошукових запитів формується набір валідних документів (рис. 4.51). Таким чином реалізується вибір документів.

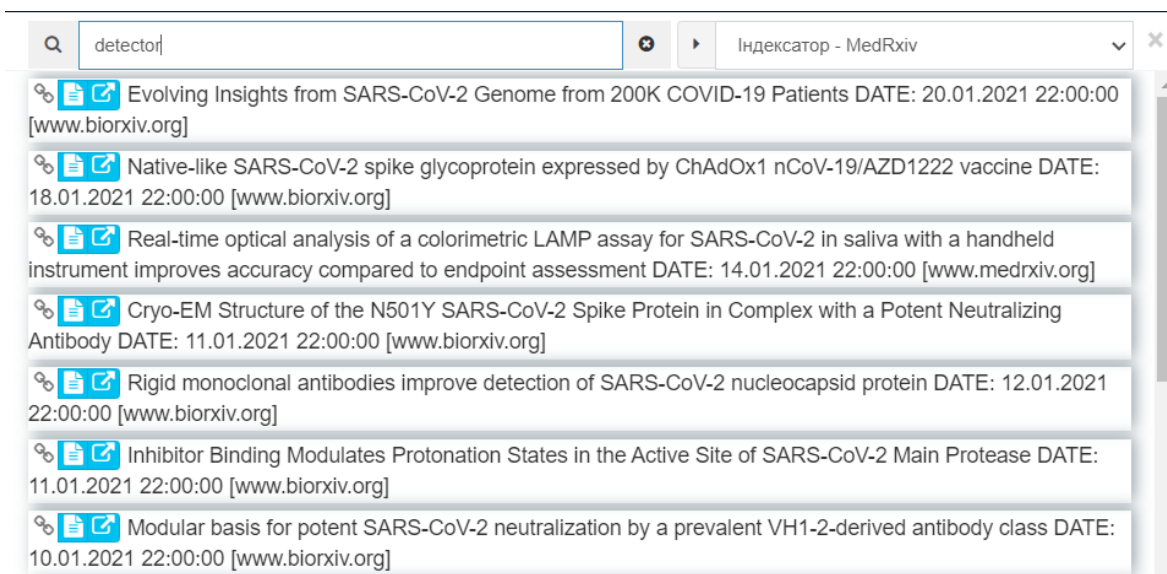


Рисунок 4.51 – Результат вибору документів для об'єкту detector

Отже, технологічною ознакою «Кластеру знань про коронавірусну інфекцію» є надання експертам можливості формулювати свої класифікаційні ознаки та умови вибору необхідної інформації у форматі термінологічних таксономій. Вказані таксономії визначають цільові контексти вибору наукової продукції, що є об'єктом пошуку та класифікації.

Використання таксономічного формату реалізує процес формування особистої бібліотеки експерта.

З докладною інструкцією користувача Кластеру можна ознайомитись за документом [151].

#### 4.7. Приклад формування критеріїв вибору інформації

Виникає проблема щодо виявлення необхідної інформації, яка знаходиться у розподілених інформаційних ресурсах, зокрема різноманітних каталогах, довідниках, збірниках, тощо. Особливо це актуально для обробки видань, які періодично оновлюються. Наприклад щорічний збірник космічного транспорту (рис. 4.52). В кожному новому випуску збірника інформація про існуючі ракетні носії оновлюється і виникає необхідність її оперативної обробки. Для цього на базі КІТ «Поліедр» створено СППР «Launch Vehicles».

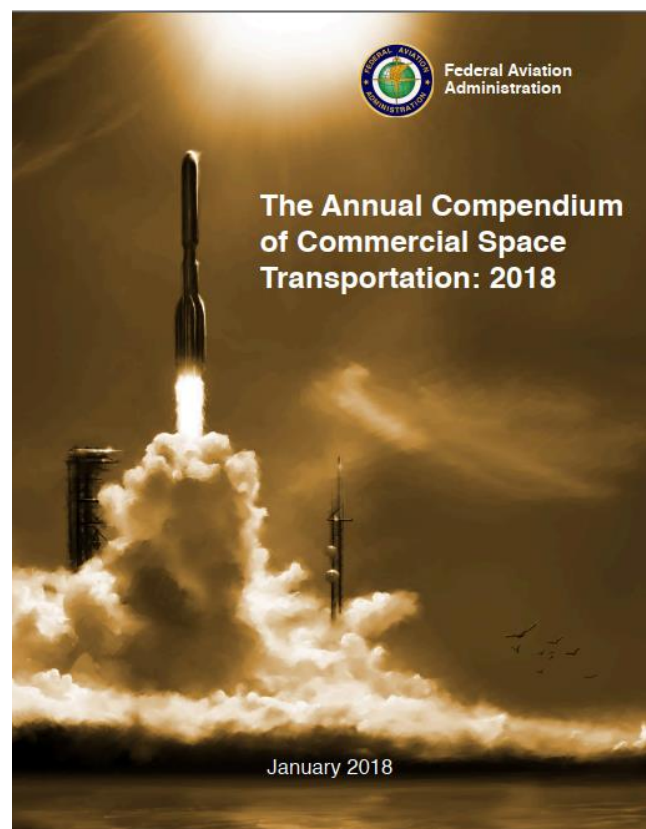
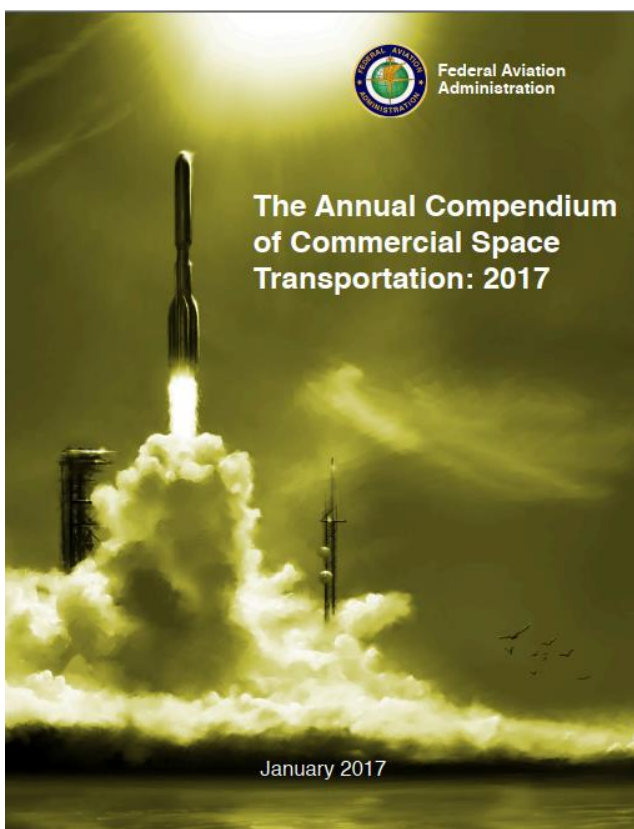
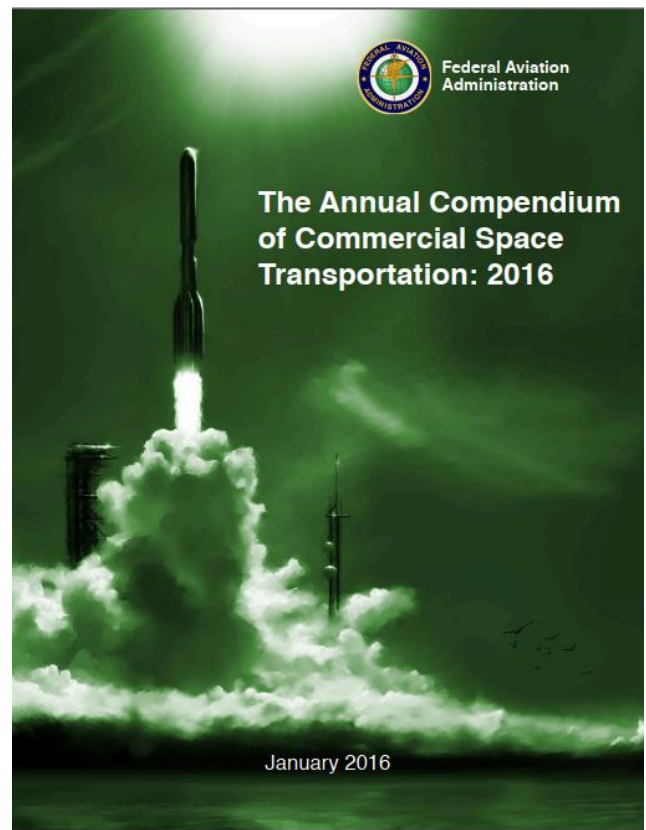
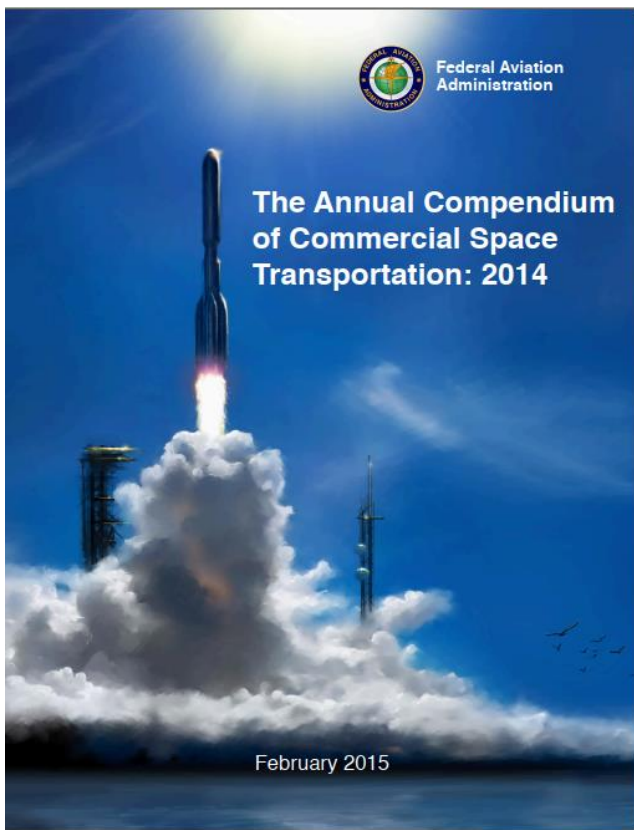


Рисунок 4.52 – Щорічний збірник космічного транспорту

СППР являє собою онтологію, де в якості об'єктів виступають моделі ракет, об'єднані в класи за країною-виробником (рис. 4.53).

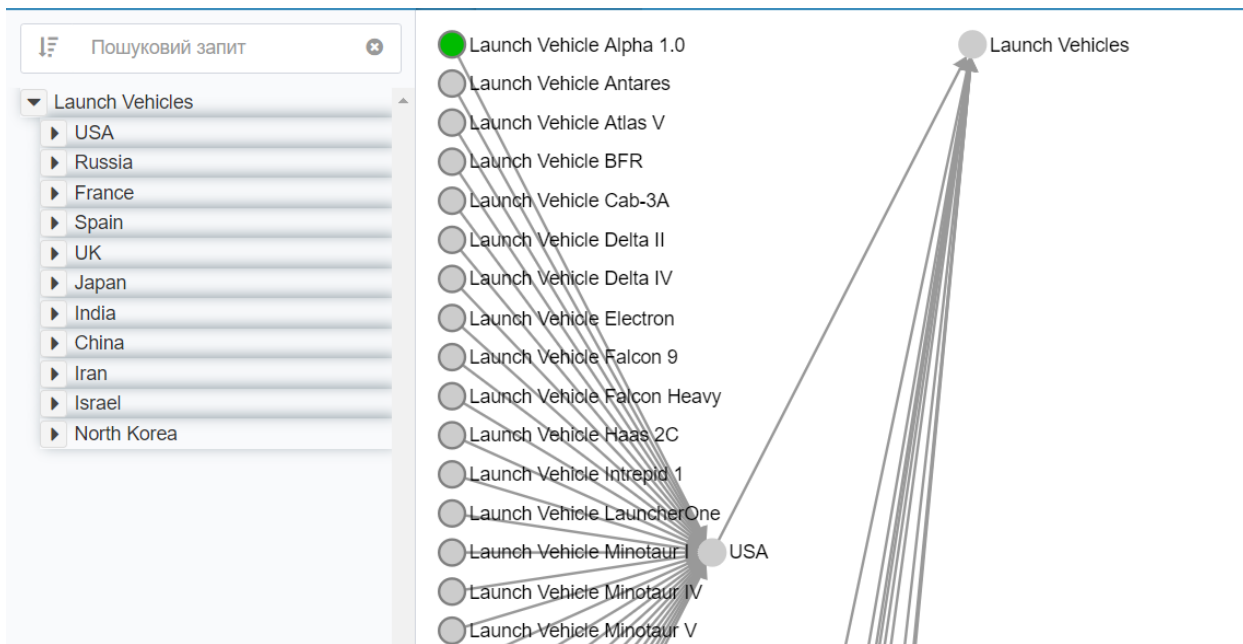


Рисунок 4.53 – Онтологія СППР «Launch Vehicles»

Кожен об'єкт має більш ніж 30 атрибутів. Значення деяких атрибутів може змінюватися з часом, що і відображається у відповідних інформаційних джерелах (збірниках). Тому важливо створити форму структурованого представлення інформації про ракети-носії з можливістю її оперативного оновлення, що і реалізовано в рамках даної СППР. Структуроване представлення інформації зображено на рис. 4.55, на відміну від слабкоструктурованого, що міститься в збірниках (рис. 4.54). Назви моделей ракет і найменування їх атрибутів є критеріями вибору інформації з наступних номерів збірника при їх автоматизованій обробці.

Даний етап структурованого представлення інформації необхідний для подальшої реалізації функцій СППР «Launch Vehicles», таких як: побудова онтологій пошукового запиту та ранжування альтернатив.

### Launch Vehicle Fact Sheet Alpha 1.0




U.S.-based company Firefly Aerospace Inc. ("Firefly") plans to offer the Alpha launch vehicle to service the burgeoning small-satellite industry. Alpha is a two-stage vehicle, utilizing composite materials to create strong lightweight primary structures such as the airframe and propellant tanks.

Led by CEO Tom Markusic and a team of space industry veterans, Firefly has recently reorganized in response to a funding shortfall in the fall of 2016. Firefly Aerospace is now majority-owned by Noosphere Ventures, the strategic venture arm of Noosphere Global, which has the resources to independently and fully fund Firefly through first launch. Firefly has re-hired much of its core team and the company is on track to achieve its launch goals. Work on the upgraded Alpha launch vehicle, scheduled for launch in mid-2019, and the conceptual design of its larger Beta vehicle, is advancing rapidly through engineering and regular engine testing.

Texas is home to both Firefly's headquarters and test facilities. A 20,000-square foot design campus in Cedar Park, Texas, just north of Austin, houses the corporate headquarters, engineering staff, prototyping facilities and machine shop. Firefly is establishing additional international offices and strategic partnerships to effectively serve the global small-satellite launch market.

<b>Launch Service Provider</b>	Firefly Aerospace, Inc.
<b>Organization Headquarters</b>	USA
<b>Manufacturer</b>	Firefly Aerospace, Inc.
<b>Mass, kg (lb)</b>	54,000 (119,050)
<b>Length, m (ft)</b>	29 (95)
<b>Diameter, m (ft)</b>	2 (6.6)
<b>Year of Planned First Launch</b>	2019
<b>Launch Site</b>	Undisclosed
<b>LEO Capacity, kg (lb)</b>	1,000 (2,205)
<b>SSO Capacity, kg (lb)</b>	650 (1,433)
<b>Estimated Price per Launch</b>	\$10M

Detailed information can be found in the Annual Compendium of Commercial Space Transportation (www.faa.gov/officeofcommercial\_space\_transportation\_reports\_and\_publications/ast). Data is maintained in FAA AST's Space Transportation Analysis and Research (STAR) database. Publication produced for FAA AST by The Tauri Group under contract.

**FAA Office of Commercial Space Transportation**  
800 Independence Avenue SW (AST)  
Washington, DC 20591  
202-267-5450 (fax)  
http://www.faa.gov/gpoast

### Launch Vehicle Fact Sheet Alpha 1.0




A 3D diagram of the four Reaver engines that will be used to power the Alpha's first stage. (Source: Firefly Aerospace)

Fairing	Length, m (ft)	Diameter, m (ft)
Standard Fairing	5 (16.4)	2 (6.6)


	1 <sup>st</sup> Stage	2 <sup>nd</sup> Stage
Stage designation	1 <sup>st</sup> Stage	2 <sup>nd</sup> Stage
Length, m (ft)	18.3 (60)	5.9 (19.4)
Diameter, m (ft)	1.8 (6)	1.8 (6)
Manufacturer	Firefly Aerospace, Inc.	Firefly Aerospace, Inc.
Propellant	LOX/Kerosene	LOX/Kerosene
Propellant mass, kg (lb)	Undisclosed	Undisclosed
Total thrust, kN (lb)	729 (164,000)	80 (18,000)
Engine(s)	4 x Reaver	1 x Lightning
Engine manufacturer	Firefly Aerospace	Firefly Aerospace
Engine thrust, kN (lb)	182 (41,000)	80 (18,000)

Detailed information can be found in the Annual Compendium of Commercial Space Transportation (www.faa.gov/officeofcommercial\_space\_transportation\_reports\_and\_publications/ast). Data is maintained in FAA AST's Space Transportation Analysis and Research (STAR) database. Publication produced for FAA AST by The Tauri Group under contract.

**FAA Office of Commercial Space Transportation**  
800 Independence Avenue SW (AST)  
Washington, DC 20591  
202-267-5450 (fax)  
http://www.faa.gov/gpoast

Рисунок 4.54 – Представлення інформації про ракети-носії в збірнику

21 21 21 Launch Vehicle Alpha 1.0



**Launch Service Provider:** Firefly Aerospace, Inc.

**Organization Headquarters:** USA

**Manufacturer:** Firefly Aerospace Inc.

**Launch Site:** -

**Estimated Price per Launch, M\$:** 10

**Year of First Launch:** 2019

**Vehicle**

**Mass, kg:** 54000

**Length, m:** 29

**Diameter, m:** 2

**Capacity, kg**

**LEO:** 1000

**SSO:** 650

21 21 21 Launch Vehicle Alpha 1.0

**Fairings**

	Length, m	Diameter, m
Standard Fairing	5	2

**Stages**

Stage name	Stage designation	Stage Length, m	Stage Diameter, m	Manufacturer	Propellant	Propellant mass, kg	Total thrust, kN
1st Stage	1st Stage	18.3	1.8	Firefly Aerospace Inc.	LOX/Kerosene	-	729
2nd Stage	2nd Stage	5.9	1.8	Firefly Aerospace Inc.	LOX/Kerosene	-	80

U.S.-based company Firefly Aerospace Inc. ("Firefly") plans to offer the Alpha launch vehicle to service the burgeoning small-satellite industry. Alpha is a two-stage vehicle, utilizing composite materials to create strong lightweight primary structures such as the airframe and propellant tanks.

Led by CEO Tom Markusic and a team of space industry veterans, Firefly has recently reorganized in response to a funding shortfall in the fall of 2016. Firefly Aerospace is now majority-owned by Noosphere Ventures, the strategic venture arm of Noosphere Global, which has the resources to independently and fully fund Firefly through first launch. Firefly has re-hired much of its core team and the company is on track to achieve its launch goals. Work on the upgraded Alpha launch vehicle, scheduled for launch in mid-2019, and the conceptual design of its larger Beta vehicle, is advancing rapidly through engineering and regular engine testing.

Texas is home to both Firefly's headquarters and test facilities. A 20,000-square foot design campus in Cedar Park, Texas, just north of Austin, houses the corporate headquarters, engineering staff, prototyping facilities and machine shop. Firefly is establishing additional international offices and strategic partnerships to effectively serve the global small-satellite launch market.

A 3D diagram of the four Reaver engines that will be used to power the Alpha's first stage. (Source: Firefly Aerospace)

Рисунок 4.55 – Представлення інформації про ракети-носії в СППР Launch Vehicles

#### **4.8. Оцінка ефективності впровадження створених онтолого-керованих засобів представлення неструктурованої інформації**

Згідно [154] головною характеристикою оцінювання ефективності застосування інформаційних систем є швидкість обробки одиниці інформації та питомі витрати на обробку одиниці інформації. Однак слід відмітити наступне: в реальних технологічних умовах існує досить великі обсяги інформаційних ресурсів, які були створені за старими технологіями і для сучасних систем є недоступними. Але сама інформація є корисною та актуальною. Такі інформаційні ресурси спеціалісти називають «темними». Зрозуміло, що реалізація процесу оброблення такої інформації суттєво підвищує ефективність довільної інформаційної системи, сервіси якої спроможні забезпечити використання такої «темної» інформації.

Якщо розглядати темні інформаційні ресурси, як певний тип представлення знань, то вирішення цієї проблеми реалізується на засадах таксономізації таких інформаційних ресурсів. Таксономізація не визначає операціональність інформаційної системи. Вона визначає тільки структурне представлення тієї інформації, яка належить системі і обробляється її сервісами. Проблема спрощено виглядає як технологічна неспроможність цих сервісів опрацьовувати інформацію, яка за форматом була сформована за вимогами її операціональності.

Тому система, яка спроможна обробити такі старі, «темні», але досить актуальні дані вже ефективна.

Розглянемо ефективність ОКЗПНІ. Спочатку за характеристикою «швидкість обробки одиниці інформації». За одиницю будемо брати поняття – документ, що використовується. Документ береться за основу тому, що його таксономічна структура змінює послідовність, оброблення та використання контекстів, які його змістовно складають.

Оцінку ефективності функціональних можливостей, у тому числі з точки зору часу, що необхідний для проведення певної операції, визначимо на засадах

застосування когнітивних функцій ОКЗПНІ. Для порівняння беруться середньо статистичні дані щодо оброблення текстових документів у системах документообігу.

Тут слід відзначити, що ефективність застосування когнітивних сервісів для формування інтерактивних документів залежить від того наскільки оперативно та якісно було проведено етапи первинного оброблення інформації. На вказані категорії оцінювання впливають певні параметри, які і визначають рівень ефективності формування онтологічної інтерактивної системи знань.

В основу об'єктивності оцінювання покладемо обсяги інформації, що обробляються та аналізуються. Для цього візьмемо множину документів, які мають у сумі 1000 сторінок (формат А4). Умовно на 1 сторінці міститься близько 300 слів. Усього – 300000 слів.

Дані щодо оброблення такого обсягу інформації наведено у табл. 4.1.

Таблиця 4.1 – Оцінка швидкості обробки документів

№ з/п	Параметри оцінювання	ОКЗПНІ	Довільна ІАС	Відносний коефіцієнт ефективності (ВКЕ)/ (пропорційність оброблення)	Абсолютний коефіцієнт ефективності / 1: ВКЕ
1.	Структуризація	Автоматизоване / 4-8 годин	Ручний режим / до тижня	8 : 40 / 0,2	5
2.	Класифікація	Автоматизоване / 4-8 годин	Ручний режим / до тижня	8 : 40 / 0,2	5
3.	Онтологічне представлення	Автоматизоване / 4-8 годин	Ручний режим тиждень	8 : 40 / 0,2	8
4.	Повнота пошуку	100 %	Відсутня індексація	100 : – / не визначимо	100
5.	Консолідація контекстів документів	Автомат індекс, кластеризація контекстів	Відсутня індексація	100 : – / не визначимо	100
6.	Зручність форми	Кластер Максимум 1 хвилина	Вихідні таблиці доба	1 : 8 / 0,125	8
7.	Повнота масиву	100% / усі контексти	100% / усі контексти	1 : 1 / 1	1
8.	Обмін інформацією	Міжконтекст 1 хвилина	Локальні обговорення Одна доба	1 : 480 / 0,002	500

## Продовження таблиці 4.1

9.	Наявність засобів оцінювання інформації	Присутні	Відсутні	8 : невизначено	100
10	Використання «темної» інформації	Макимум один тиждень	Неспроможна	40 : – / не визначено	100
					927 / середнє значення – 92

Критерії, які у процесі оцінювання не визначено, будемо вважати такими, що дорівнюють нулю «0». Абсолютний коефіцієнт має значення сто «100». Тоді середнє значення коефіцієнту ефективності ОКЗПНІ, відносно семантичному аналізу документів становить 92. Тобто умовна одиниця інформації – документ, обробляється когнітивними сервісами ОКЗПНІ, в 92 рази швидше.

Однак треба враховувати, що сервіси ОКЗПНІ реалізують використання «темних» даних, на що сервіси систем документообігу взагалі неспроможні. Отже, ефективність ОКЗПНІ піднімається фактично експоненціально.

#### 4.9. Висновки за розділом 4

Розроблено інструментальний засіб «Система автоматизованої обробки різностильових документів», який надає можливість витягу із множини текстових документів знань, релевантних до заданої предметної галузі, їх системно-онтологічної структуризації та формально-логічного подання.

Проілюстровано застосування науково-технологічних та методичних засобів онтолого-керованого трансдисциплінарного представлення великих масивів слабо і неструктурованої інформації, для підтримки процесів пошуку, класифікації та вибору необхідного контенту. Створено знання-орієнтовані персональні бібліотеки та репозиторії. Виявлено критеріальні ознаки щодо оцінювання та прийняття рішень на основі методології системно-онтологічного аналізу інформаційних ресурсів.



Проведено експериментальні дослідження в ході яких підтверджено ефективність використання розробленої системи.

На основі концептуальної, інформаційної моделей та інструментальних засобів представлення неструктурованої інформації в онтологічному форматі фактично визначено метод трансдисциплінарного представлення природномовних видань.

Основні результати досліджень для даного розділу опубліковано в [10, 149, 152, 153].

## ВИСНОВКИ

В дисертаційному дослідженні вирішена актуальна науково-технічна задача систематизації різностильових природномовних видань в глобальному середовищі шляхом розробки моделі, методу і засобів їх трансдисциплінарного представлення та консолідації.

В процесі виконання дисертаційної роботи отримано такі основні результати:

Проведено огляд існуючих засобів для обробки та представлення природномовних текстів і виявлено нестачу інструментальних засобів, здатних представляти результати обробки неструктурованих текстових масивів у вигляді, що задовольняв би критерію трансдисциплінарності.

Розроблено концептуальну онтолого-керовану модель трансдисциплінарного представлення природномовних видань в електронному вигляді.

Розроблено метод трансдисциплінарного представлення природномовних видань, що описує функціональну архітектуру трансдисциплінарного образу документу та процедуру формування онтологічного опису нормативного документу на основі його текстового представлення.

Створено мережевий програмно-інформаційні засіб трансдисциплінарного представлення інформації шляхом формування онтологій видань на основі репрезентації їх таксономій, зв'язків між їх контекстами та множинного представлення видань, що забезпечує ефективний пошук інформації у природномовних виданнях.

Набув подальшого розвитку метод рекурсивної редукції шляхом побудови перетворення, яке приводить до предикативної форми онтологічно задані описи інформації.

Удосконалено технологічні аспекти об'єктно-орієнтованого підходу до пошуку інформації.

Розроблено трансдисциплінарний метод представлення природномовних видань, який базується на концептуальній та інформаційній моделях представлення неструктурованої інформації в онтологічному форматі; класифікації

різностильових документів; шкалі ступеня формалізованості документу; інструментальному засобу «Рекурсивний редуктор»; процедурі формування онтологічного опису системи документів на основі його текстового представлення; правилах автоматизованого виділення інформації з документів.

Створено онтологічний репозиторій нормативних документів.

Набули подальшого розвитку ідеї ефективної структуризації та, відповідно, швидкого пошуку інформації.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы. М. : Академия. 2006. 48 с.
2. Jurafsky, D., Martin G. H. *Speech and Language Processing*. Prentice Hall. 2000. URL: <https://web.stanford.edu/~jurafsky/slp3/> (дата звернення: 10.05.2021).
3. “Обробка природної мови”. URL: [https://uk.wikipedia.org/wiki/Обробка\\_природної\\_мови](https://uk.wikipedia.org/wiki/Обробка_природної_мови) (дата звернення: 17.06.2021).
4. Jones K. S. *Natural language processing: a historical review*. 1994. URL: <https://aclanthology.org/www.mt-archive.info/Zampolli-1994-Sparck-Jones.pdf> (дата звернення: 04.11.2022).
5. Liddy E. D. *Natural Language Processing. Encyclopedia of Library and Information Science*. NY : Marcel Decker, Inc. 2001.
6. Barkovska, O., Khomych, V., Nastenka O. Дослідження методів обробки та аналізу тексту при організації електронних сховищ інформаційних об’єктів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 1 (19). С. 5–12. DOI: <https://doi.org/https://doi.org/10.30837/ITSSI.2022.19.005>
7. Praveena, S., Justus S. A Study on Knowledge Representation Models. *European Journal of Molecular and Clinical Medicine*. 2020. вип. 7. № 4. С. 2446–2452.
8. Комарницька О. Методи автоматизованого семантичного аналізу природномовної інформації. *Філологічний дискурс*. 2018. № 7. С. 92–100.

9. Палагин, А. В., Крытый, С. Л., Петренко М. Г. Формализация проблемы извлечения знаний из естественных языковых текстов. *Information Technologies & Knowledge*. 2012. вып. 6. № 3. С. 203–218.
10. Гайко С. І. Онтолого - керовані засоби оброблення та подання великих масивів неструктурованих текстів. 2023. вип. 2. № 24. С. 27–38. DOI: <https://doi.org/https://doi.org/10.30837/ITSSI.2023.24.027>
11. Хомский Н. Синтаксические структуры. *Новое в лингвистике*. 1962. № II.
12. Кудрявцев, В. Б., Алешин, С. В., Подколзин А. С. Введение в теорию автоматов. М. : Наука. 1985.
13. Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. М. : Мир. 1978.
14. Вудс В.А. Сетевые грамматики для анализа естественных языков. *Кибернетический сборник. Новая серия*. 1978. № 13. С. 120–158.
15. Sleator, D., Temperley D. Parsing English with a Link Grammar. *Third International Workshop on Parsing Technologies*. 1993. С. 277–292.
16. Бухараев Р. Г. Основы теории вероятностных автоматов. М. : Наука. 1985.
17. Мельчук И. А. Опыт теории лингвистических моделей «Смысл ↔ Текст». М. : Наука. 1974.
18. Диковицкий В. В., Шишаев, М. Г. Обработка текстов естественного языка в моделях поисковых систем. *Труды Кольского научного центра РАН*. 2010. С. 29–34.
19. Shannon C. E. Prediction and Entropy of Printed English. *Bell System Technical Journal*. 1951. вып. 30. № 1. С. 50–64. DOI: <https://doi.org/https://doi.org/10.1002/j.1538-7305.1951.tb01366.x>
20. Кузин Л.П. Основы кибернетики: в 2-х т. 1979.
21. Broder A Identifying and Filtering Near-Duplicate Documents. *11th Annual Symposium on Combinatorial Pattern Matching*. 2000. С. 1–10.
22. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа. М. : ООО «Вильямс». 2005. 275 с.

23. Settles B. Biomedical named entity recognition using conditional random fields and rich feature sets. *International Joint Workshop on Natural Language Processing in Biomedicine and its Applications*. 2004. С. 104–107.
24. Huang X., Jack M., Ariki Y. *Hidden Markov Models for Speech Recognition*. Edinburgh University Press. 1990.
25. Бабин Д. Н., Холоденко А. Б. Об автоматной аппроксимации естественных языков. *Интеллектуальные системы*. 2008. вип. 12. № 1–4. С. 125–136.
26. Савченко Д. С. Теоретичні основи формування моделей систем автоматизованої обробки неструктурованих текстів. *Інформаційна безпека людини, суспільства, держави*. 2015. вип. 2. № 18. С. 50–59.
27. Parreiras F. S. *Semantic Web and Model-Driven Engineering*. IEEE Press. 2012. 272 с.
28. Субботін С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень. Запоріжжя : ЗНТУ. 2008. 341 с.
29. Синиця А. С. Прагматико-когнітивна інтерпретація сучасної аналітичної філософії: дис. докт. філос. наук. Львівський національний університет ім. І. Франка. 2018. 453 с.
30. Палагин, А. В., Крытый, С. Л., Петренко М. Г. *Онтологические методы и средства обработки предметных знаний*. Луганск : изд-во ВНУ им. В. Даля. 2012. 324 с.
31. Yang, L., Cormican, K., Yu M. *Ontology-based systems engineering: A state-of-the-art review*. *Computers in Industry*. 2019. вип. 111. С. 148–171. DOI: <https://doi.org/https://doi.org/10.1016/j.compind.2019.05.003>
32. Глушков В. М. *Обработка информационных массивов в автоматизированных системах управления*. Киев : Наукова думка. 1970. 183 с.
33. Глушков В. М. Шляхи створення штучного інтелекту. *Наука і культура*. 1981. № 3. С. 130–136.
34. Гладун В. П. *Процессы формирования новых знаний*. София : СД «Педагог 6». 1994. 189 с.

35. Святогор Л. А., Гладун В. П. Машинное понимание текстов естественного языка : онтологическая парадигма. *Штучний інтелект*. 2010. № 3. С. 249–258.
36. Широков В. А., Рабулець О. Г. Формалізація в галузі лінгвістики. *Актуальні проблеми української лінгвістики: теорія і практика*. 2002. № 5. С. 3–28.
37. Широков В. А. Язык. Информация. Система : Трансдисциплинарность в лингвистике. Киев. 2017. 280 с.
38. Хорошевский В. Ф. Программные средства представления знаний. *Искусственный интеллект*. М : Радио и связь. 1990. С. 72–82.
39. Поспелов Д. А. Представление знаний. Опыт системного анализа. *Системные исследования. Методологические проблемы*. 1985. № 17. С. 83–102.
40. Овдій О. М. До питання автоматизації проектування робочих процесів на основі алгебро-алгоритмічного та онтологічного інструментарію. *Проблеми програмування*. 2019. № 1. С. 37–47.
41. Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И. Инженерия знаний. Модели и методы. СПб. 2016.
42. Gavrilova T., Kokoulina L. Ontology Engineering to Design Artificial Intelligence Course. *In Smart Education and e-Learning SEEL, Smart Innovation, Systems and Technologies. Springer*. 2019. вип. 144. С. 201–207.
43. Загорулько Ю. А. Построение порталов научных знаний на основе онтологий. *Вычислительные технологии*. 2007. вип. 12. № 2. С. 169–177.
44. Malyshevsky A. Qualitative models in the theory of complex systems. Science. Fizmatlit. 1998. 528 с.
45. Андон П., Дерещкий В. Проблеми побудови сервіс-орієнтованих прикладних інформаційних систем в semantic web середовищі на основі агентного підходу. *Проблеми програмування*. 2006. № 2–3. С. 493–502.
46. Валькман Ю. Р., Рыхальский А. Ю. Модельно-параметрическое пространство – средство представления знаний исследователей сложных систем. *УСiМ*. 2009. № 1. С. 20–30.
47. Стрижак О. Є. Онтологічні інформаційно-аналітичні системи. *Комп'ютерні*

*системи та інформаційні технології*. 2014. № 3. С. 71–76.

48. Стрижак О. Є. Управління знаннями - головна парадигма сучасної освіти. *Комп'ютер у школі та сім'ї*. 2016. № 5. С. 9–11.
49. Стрижак О. Є. Засоби онтологічної інтеграції і супроводу розподілених просторових та семантичних інформаційних ресурсів. *Екологічна безпека та природокористування*. 2013. № 1. С. 166–177.
50. Величко В. Ю., Малахов К. С., Семенов В. В. С. А. Е. Комплексные инструментальные средства инженерии онтологий. *Information Models and Analyses*. 2014. вип. 3. № 4. С. 336–361.
51. Gruber T. A translation approach to portable ontology specifications. *Knowledge Acquisition*. 1993. вип. 5. № 2. С. 199–220.
52. Guarino N. Formal ontology and information systems. *First international conference (FOIS'98)*. IOS press. 1998.
53. Borgo S., Guarino N. V. L. Formal ontology for semanticists. *Research Institute for Computer Science of Toulouse*. 2005. С. 3–15.
54. Noy N. Semantic Integration : A Survey Of Ontology-Based Approaches. *ACM Sigmod Record*. 2004. вип. 33. № 4. С. 65–70.
55. Gomez-Perez A., Fernandez-Lopez M., Corcho O. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Verlag. 2004.
56. Uschold M., Gruninger M. *Ontologies : Principles , Methods and Applications*. *The knowledge engineering review*. 1996. вип. 11. № 2. С. 93–136.
57. Happel H. S. S. Applications of ontologies in software engineering. 2006. URL: [https://km.aifb.kit.edu/ws/swese2006/final/happel\\_full.pdf](https://km.aifb.kit.edu/ws/swese2006/final/happel_full.pdf) (дата звернення: 10.05.2021).
58. Lassila O., Mc Guinness D. The role of frame-based representation on the semantic web. *Linköping Electronic Articles in Computer and Information Science*. 2001. вип. 6. № 5.
59. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем.

СПб : Питер. 2001. 384 с.

60. Осуга С. Обработка знаний. М. : Мир. 1990. 293 с.
61. Озадовська Л. Єдності знань принцип. *Філософський енциклопедичний словник. Абрис*. 2002. С. 742.
62. Тишко О. В., Коцюк Л. М. Лінгвістичний аналіз поняття фрейму. *Наукові записки НУ «Острозька академія»*. 2009. № 11. С. 391–399.
63. Методи представлення знань. URL: <https://uk.wikipedia.org/wiki?curid=2437512> (дата звернення: 10.05.2021).
64. Джарратано Д. Представление знаний. *Экспертные системы: принципы разработки и программирование*. М. : Издательский дом «Вильямс». 2006. С. 1152.
65. Вудс У. А. Основные проблемы представления знаний. *ТИИЭР*. 1986. вып. 74. № 10. С. 32–47.
66. Klein J. T. *Transdisciplinarity: Joint Problem Solving Among Science, Technology, and Society: An Effective Way for Managing Complexity*. Birkhäuser. 2001.
67. Nicolescu B. *Transdisciplinarity - Theory and Practice*. Hampton Press, Cresskill, NJ, USA. 2008. 320 с.
68. Judge A. Conference Paper. *1st World Congress of Transdisciplinarity, Union of International Associations*. 1994.
69. Князева Е. Н. Трансдисциплинарные стратегии исследований. *Вестник ТГПУ*. 2011. № 10. С. 193–201.
70. Киященко Л. П., Гребенщикова Е. Г. Современная философия науки: трансдисциплинарные аспекты. М. : МГМСУ. 2011. 172 с.
71. Lawrence, M., Williams, S., Nanz, P., Renn O. Characteristics, potentials, and challenges of transdisciplinary research. *One Earth*. 2022. вып. 5. № 1. С. 44–61. DOI: <https://doi.org/https://doi.org/10.1016/j.oneear.2021.12.010>
72. Renn O. Transdisciplinarity: Synthesis towards a modular approach. *Futures*. 2021. вып. 130. С. 1–18. DOI: <https://doi.org/https://doi.org/10.1016/j.futures.2021.102744>



73. Гуреев П. М. Современная наука и методология трансдисциплинарности. *Вестник университета*. 2013. № 1. С. 172–180.
74. Стрижак О. Є. Трансдисциплінарна інтеграція інформаційних ресурсів: дис. д-ра техн. наук. Ін-т телекомунікацій і глобал. інформ. простору. 2014. 563 с.
75. Палагин А. В. Трансдисциплинарность, информатика и развитие современной цивилизации. *Вістник НАН України*. 2014. № 7. С. 25–33.
76. Gruber T. What is an Ontology? URL: <https://web.archive.org/web/20100716004426/http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (дата звернення: 10.05.2021).
77. Басюк, Т. М., Досин, Д. Г., Литвин В. В. Онтологічний інжиніринг. Львів : Вид-во Львівської політехніки. 2017. 224 с.
78. Добров Б. В., Иванов В. В., Лукашевич Н. В., Соловьев В. Д. Онтологии и тезаурусы: модели, инструменты, приложения: учебное пособие. М : БИНОМ. 2013. 174 с.
79. Стрижак А. Е., Трофимчук А. Н., Цурика Л. Ю. Трансдисциплинарные онтологии – информационная платформа проведения экологических экспертиз. *Екологічна безпека та природокористування*. 2014. № 16. С. 128–137.
80. Стрижак А. Е. Онтологические аспекты трансдисциплинарной интеграции информационных ресурсов. *Открытые информационные и компьютерные интегрированные технологии*. 2014. № 65. С. 211–223.
81. Когаловский М. Р. Перспективные технологии информационных систем. М. : ДМК Пресс. 2018. 287 с.
82. Горборуков В. В. Технологічні засоби онтологічного супроводу розв'язання задач ранжування альтернатив: дис. канд. техн. наук. Інститут телекомунікацій і глобального інформаційного простору НАН України. 2018. 142 с.
83. Гайко С., Приходнюк В. Средства трансдисциплинарного представления информационных ресурсов разных стилей. *Information Models and Analysis*. Sofia : ITNEA. 2020. вип. 9. № 1. С. 78–99.
84. Гайко, С. І., Приходнюк В. В. Підхід до автоматизованої структуризації освітніх

ресурсів на основі методу рекурсивної редукції. *Наукові записки Малої академії наук України*. 2021. № 1 (20). С. 28–38. DOI: <https://doi.org/http://doi.org/10.51707/2618-0529-2021-20-03>

85. Величко В. Ю., Попова М. А., Приходнюк В. В., Стрижак О. Є. ТОДОС – ІТ-платформа формування трансдисциплінарних інформаційних середовищ. *Системи озброєння і військова техніка*. 2017. № 1. С. 10–19.
86. Довгий С. О., Величко В. Ю., Глоба Л. С., Стрижак О. Є., Андрущенко Т. І. та ін Комп'ютерні онтології та їх використання у навчальному процесі. 2013. 310 с.
87. Кудрявцев Д. В. Системы управления знаниями и применение онтологий. СПб : Изд-во Политехн. ун-та. 2010. 344 с.
88. Попова М. А. Методика формування та використання комп'ютерних онтологій в галузі екологічної освіти. К. : СІТШПІНТ. 2013. 200 с.
89. Малишевский А. В. Качественные модели в теории сложных систем. М. : Наука. Физматлит. 1998. 528 с.
90. Стрижак О. Є., Приходнюк В. В., Гайко С. І., Шаповалов В. Б. Відображення мережевої інформації у вигляді інтерактивних документів. Трансдисциплінарний підхід. *Математичне моделювання в економіці*. 2018. № 3. С. 87–100.
91. Володченко, А., Стрижак, О., Храпач Г. Трансдисциплінарний характер операціональності розвитку обдарованості учнівської молоді. *Навчання і виховання обдарованої дитини: теорія і практика*. 2016. № 1. С. 100–110.
92. Подлипаев В. А., Стрижак А. Е. Интеграция информационных ресурсов различной природы в сетевентрической среде на основе категории трансдисциплинарности. *Системи озброєння і військова техніка*. 2018. № 3(55). С. 85–94. DOI: <https://doi.org/10.30748/soivt.2018.55.12>
93. Гончар, А. В., Стрижак, О. Є., Беркман Л. Н. Трансдисциплінарна консолідація інформаційних середовищ. *Зв'язок*. 2021. № 1(149). С. 3–10. DOI: <https://doi.org/10.31673/2412-9070.2021.010310>

94. Stryzhak O., Horborukov V., Prychodniuk V., Franchuk O. C. R. Decision-making System Based on The Ontology of The Choice Problem. *International Symposium on Automation, Information and Computing*. 2021. вип. 1828. № 1. DOI: <https://doi.org/>. DOI: <https://doi.org/10.1088/1742-6596/1828/1/012007>
95. Артемьева И. Л., Рештаненко Н. В. Интеллектуальная система, основанная на многоуровневой онтологии химии. *Программные продукты и системы*. 2008. № 1. С. 84–87.
96. Боровикова О. И., Загорулько Ю. А. Организация порталов знаний на основе онтологий. *Диалог '2002 : Труды международного семинара*. Протвино. 2002. С. 76–82.
97. Гладун А. Я., Рогущина Ю. В. Онтологии в корпоративных системах. *Корпоративные системы*. 2006. вип. 1.
98. Горборуков В. В., Стрижак О. Є., Франчук О. В. Використання онтологій у системах підтримки прийняття рішень. *Математичне моделювання в економіці*. 2013. № 3. С. 33–39.
99. Гуржій А. М., Стрижак О. Є. Онтологічні інструменти управління мережевими інформаційними ресурсами та їх використання в освітній та науковій діяльності. *Наукові записки Малої академії наук України*. 2013. № 2. С. 427–434.
100. Загорулько Ю. А. Автоматизация сбора онтологической информации об интернет-ресурсах для портала научных знаний. *Известия Томского политехнического университета*. Управление, вычислительная техника и информатика. 2008. С. 114–119.
101. Палагин А. В., Петренко Н. Г. К вопросу системно-онтологической интеграции знаний предметной области. *Математические машины и системы*. 2007. № 3,4. С. 63–75.
102. Палагин А. В., Яковлев Ю. С. Системная интеграция средств компьютерной техники. Винница : УНІВЕРСУМ. 2005. 680 с.
103. Палагин А. В., Риппа С. П., Саченко А. А. Концептуализация и проблематика

- онтологий. *Искусственный интеллект*. 2008. № 3. С. 374–379.
104. Палагин А. В., Крытый С. Л., Величко В. Ю., Петренко Н. Г. К анализу естественно-языковых объектов. *Information technologies and knowledge*. ITNEA. 2009. № 3. С. 36–43.
105. Палагин А. В., Петренко Н. Г. К проектированию онтологоуправляемой информационной системы с обработкой естественно-языковых объектов. *Математические машины и системы*. 2008. № 2. С. 14–23.
106. Палагин А. В., Петренко Н. Г. Системно-онтологический анализ предметной области. *УСиМ*. 2009. № 4. С. 3–14.
107. Клещев А. С., Шалфеева Е. А. Классификация свойств онтологий. Онтологии и их классификации. Владивосток : ИАПУ ДВО РАН. 2005. 19 с.
108. Котурова М. П. Стилистика научной речи. М. : Академия. 2010. 240 с.
109. Приходнюк В. В. Технологічні засоби трансдисциплінарного представлення геопросторової інформації: дис. канд. техн. наук. Інститут телекомунікацій і глобального інформаційного простору. 2017. 267 с.
110. Takashima A., Bakker I., van Hell J.G., Janzen G. M. J. M. Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and Language*. 2017. № 167. С. 44–60. DOI: <https://doi.org/10.1016/j.bandl.2016.05.009>
111. Lamberts, K., Shanks D. Knowledge concepts and categories. London : Psychology Press. 1997. 480 с. DOI: <https://doi.org/10.4324/9780203765418>
112. Стрижак, О. Є., Володченко А. Є. Онтологія девіантної поведінки підлітків. *Педагогічні інновації: ідеї, реалії, перспективи*. 2016. вип. 1. № 16. С. 98–105.
113. Kleene S. C. Introduction to Metamathematics (Bibliotheca Mathematica). North-Holland Hardcover. 1980. 560 с.
114. Working paper NATO № AC/322-D(2019)0034 (INV) «Consultation, command and control board: C3 taxonomy baseline 3.1». North Macedonia. 2019. URL: [https://www.nato.int/nato\\_static\\_fl2014/assets/pdf/pdf\\_2019\\_09/20190912\\_190912-C3-Taxonomy-baseline.pdf](https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2019_09/20190912_190912-C3-Taxonomy-baseline.pdf) (дата звернення: 10.05.2021).

115. Dovgyi S., Stryzhak O. Transdisciplinary Fundamentals of Information-Analytical Activity. *Advances in Information and Communication Technology and Systems. MCT 2019. Lecture Notes in Networks and Systems*. 2019. вип. 152. DOI: [https://doi.org/https://doi.org/10.1007/978-3-030-58359-0\\_7](https://doi.org/https://doi.org/10.1007/978-3-030-58359-0_7)
116. Шаталкин А. И. Таксономия. Основания, принципы и правила. 2012. 600 с.
117. Палагин А. В. Онтологическая концепция информатизации научных исследований. *Кибернетика и системный анализ*. 2016. вип. 52. № 1. С. 3–9.
118. Гладун, В. П., Ващенко, Н. Д., Величко, В. Ю., Ткаченко Ю. Г. Структуризация и анализ данных в растущих пирамидальных сетях. *System Research & Information Technologies*. 2004. № 1. С. 82–92. DOI: <https://doi.org/10.1017/SBO9781107415324.004>
119. Gladun, V., Vashchenko, N., Velichko, V., Tkachenko Y. Structurization and data analysis in growing pyramidal networks. *Systems research and information technology*. 2004. № 1. С. 82–92.
120. Величко В. Логико-лингвистические модели как технологическая основа интерактивных баз знаний. *Information Models and Analyses*. 2019. вип. 8. № 4. С. 325–340.
121. Величко В. Ю. Алгоритм побудови зростаючих пірамідальних мереж у паралельному обчислювальному середовищі. *Комп'ютерні засоби мережі та системи*. 2011. № 10. С. 50–57.
122. Леонтьев А. Н. Деятельность. Сознание. Личность. М. 2006.
123. Гончар А. В. Онтологія трансдисциплінарної консолідації 3D-панорам: дис. канд. техн. наук. Інститут телекомунікацій і глобального інформаційного простору. 2021. 168 с.
124. Стрижак, О. Є., Потапов, Г. М., Приходнюк, В. В., Чепков Р. І. Еволюція управління – від ситуаційного до трансдисциплінарного. *Екологічна безпека та природокористування*. 2019. № 2 (30). С. 91–112.
125. Elson D. K. Modeling Narrative Discourse. Columbia University New York City. 2012. 383 с.

126. Барендрегт Х. Лямбда-исчисление. Его синтаксис и семантика. М. : Мир. 1985. 606 с.
127. Величко В. Ю. Науково-технологічні основи знання-орієнтованої обробки природномовних текстів та її застосування: автореф. дис. докт. техн. наук. Ін-т кібернетики ім. В. М. Глушкова. 2021. 44 с.
128. Гуржій А., Гайко С., Попова М., Стрижак О. Трансдисциплінарні когнітивні засоби підтримки наукових досліджень. *Освіта для миру=Edukacja dla pokoju*. 2019. С. 190–200.
129. Палагін, О. В., Петренко, М. Г., Величко, В. Ю., Малахов, К. С., Карун О. В. Основи проектування та розробки програмних моделей онтологокерованих комп'ютерних систем. С. 6. URL: <https://arxiv.org/ftp/arxiv/papers/1802/1802.06829.pdf> (дата звернення: 10.05.2021).
130. Городиловська Г. Проблема стилів в українському мовознавстві. *Вісник Львівського університету. Серія журналістика*. 2003. № 23. С. 136–143.
131. Арешенков Ю. Класифікація функціональних стилів і вивчення стилістики у вищій та середній школі. *Мовознавство*. 1993. № 1.
132. Гайко С. І., Приходнюк В. В. Автоматизована обробка академічних видань засобами системи «Рекурсивний редуктор». *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях*. Київ. 2018. С. 126–128.
133. Ленець К. Епістолярний стиль. *Стиль і час. Хрестоматія*. 1983.
134. Приходнюк В. В., Стрижак О. Є., Гайко С. І., Чепков Р. І. Інформаційно-аналітичний комплекс підтримки процесів трансдисциплінарних досліджень. *Екологічна безпека та природокористування*. 2018. № 4 (28). С. 103–119. DOI: <https://doi.org/10.32347/2411-4049.2018.4.103-119>
135. Дзюбишина-Мельник Н. Ще один стиль української літературної мови. *Культура слова*. 1994. вип. 45.
136. Гайко С. І., Приходнюк В. В. Застосування методу рекурсивної редукції для

- обробки великих масивів різномірних документів. *Управління знаннями та конкурентна розвідка*. Харків : ХНУРЕ. 2020. С. 181–182.
137. Мацько, Л. І., Сидоренко, О. М., Мацько О. М. Стилїстика української мови. 2003. 462 с.
138. Гайко С. І., Приходнюк В. В. Представлення освітніх ресурсів у вигляді інтерактивних документів. *Інтернет – Освіта – Наука*. Вінниця : ВНТУ. 2020. С. 249–251.
139. Непийвода Н. Мова української науково-технічної літератури (функціонально-стилістичний аспект). 1997. 303 с.
140. Гайко С. І., Зотова Л. В. Структуризація інформаційних ресурсів різних стилів у форматі наративного дискурсу. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях*. Київ. 2020. С. 175–177.
141. Дудик П. С. Стилїстика української мови: Навчальний посібник. 2005. 368 с.
142. Stryzhak O., Prykhodniuk V., Popova M., Nadutenko M., Haiko S., Cherkov R. Development of an Oceanographic Databank Based on Ontological Interactive Documents. *Lecture Notes in Networks and Systems*. Cham : Springer. 2021. вип. 284. С. 97–114. DOI: [https://doi.org/https://doi.org/10.1007/978-3-030-80126-7\\_8](https://doi.org/https://doi.org/10.1007/978-3-030-80126-7_8)
143. Гайко, С. І., Стрижак О. Є. Модель трансдисциплінарної інтеграції інформаційних ресурсів різних стилів. *Ідеї академіка В.М. Глушкова і сучасні проблеми теоретичної кібернетики*. 2020. С. 23–26.
144. Broy M. *Mathematical System Models as a Basis of Software Engineering*. Berlin : Springer. 1995. 292 с.
145. Петрик М. Р., Петрик О. Ю. Моделювання програмного забезпечення. Тернопіль : Вид-во ТНТУ імені Івана Пулюя. 2015. 200 с.
146. Стрижак О. Є., Величко В. Ю., Приходнюк В. В., Надутенко М. В., Горбурков В. В., Франчук О. В., Довгий С. О., Лісовий О. В., Палагін О. В., Сергієко І. В., Широков В. А., Романенков Ю. О., Шостак І. В., Потапов Г. М., Філістєєв Д. А., Чепков І. Б., Гупало А. Ю., Глоба Л. С., Гладун О. В., Попова М. А., Шевченко

- Л. Л., Малахов К. С. Свідоцтво про авторське право на твір №96078 від 17.02.2020 “Комп’ютерна програма “Когнітивна ІТ платформа ПОЛІЕДР” (“КІТ ПОЛІЕДР”) (“POLYHEDRON”). С. Офіційний бюлетень №57 (31.03.2020) 402-403.
147. Booch G., Rumbaugh J., Jacobson I. The Unified Modelling Language - User Guide. Addison Wesley Longman, Inc. 1999. 185 с.
148. Marshall C. Enterprise Modelling with UML. Addison-Wesley Professional. 2000. 288 с.
149. Гайко С. І. Застосування системи автоматизованої обробки різностильових документів на прикладі океанографічних законодавчих актів. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: тенденції 2020 року*. Київ. 2021. С. 203–206.
150. Інструкція користувача КІТ «Поліедр». 2020. URL: [https://storage.ulif.org.ua/storage/instructions/polyhedron\\_instruction.pdf](https://storage.ulif.org.ua/storage/instructions/polyhedron_instruction.pdf) (дата звернення: 15.02.2023).
151. COVID-19: Трансдисциплінарний кластер знань. Інструкція користувача. 2020. URL: [https://storage.ulif.org.ua/storage/COVID-19/instructions/instruction\\_covid.pdf](https://storage.ulif.org.ua/storage/COVID-19/instructions/instruction_covid.pdf) (дата звернення: 15.02.2023).
152. Гайко С. І., Приходнюк В. В. Обробка інформації в середовищі «ГОДОС-Редуктор» на прикладі тексту публіцистичного стилю. *Управління знаннями та конкурентна розвідка*. Харків : ХНУРЕ. 2019. С. 11–12.
153. Гайко С. І. Методика застосування системи «Рекурсивний редуктор». *Сучасні технології в науці та освіті*. Сєверодонецьк. 2019. С. 90–92.
154. William S. Davis, David. C. Yen. The Information System Consultant’s Handbook. Systems Analysis and Design. CRC Press. 1998. 800 с.



# ДОДАТКИ

## Додаток А. Акти впровадженнь

Затверджую  
 “ 25 ” 01 2022 р.  
 Директор Українського мовно-інформаційного фонду НАН України  
 академік НАН України  
 В.А. Широков



### АКТ

#### впровадження науково-технічної продукції

1. **Найменування пропозиції для впровадження:** Науково-технологічні та методичні засоби онтолого-керованого трансдисциплінарного представлення великих масивів слабо- і неструктурованої інформації, щодо підтримки процесів пошуку, класифікації та вибору необхідного контенту. Створення знання-орієнтованих персональних бібліотек та репозиторіїв. Виявлення критеріальних ознак щодо оцінювання та прийняття рішень на основі методології системно-онтологічного аналізу інформаційних ресурсів.
2. **Ким запропоновано, адреса:** відділ інформаційної безпеки Інституту телекомунікацій і глобального інформаційного простору НАН України; адреса: Чоколівський б-р, 13, 03186, м. Київ.
3. **Виконавець:** ГАЙКО Світлана Іванівна.
4. **Джерело інформації:**  
 Стрижак, О.Є., Приходнюк В.В., **Гайко С. І.**, Шаповалов В.Б. Відображення мережевої інформації у вигляді інтерактивних документів. Трансдисциплінарний підхід. *Математичні моделі в економіці*. 2018. № 3 (12). С. 87-100.  
 Приходнюк В.В., Стрижак О.Є., **Гайко С. І.**, Чепков Р.І. Інформаційно-аналітичний комплекс підтримки процесів трансдисциплінарних досліджень. *Екологічна безпека та природокористування*. 2018. № 4 (28). С. 103-119.  
**Гайко С. І.**, Приходнюк В. В. Автоматизовна структуризація освітніх ресурсів. *Наукові записки МАНУ*. 2021. № 1.  
**Гайко С. І.**, Приходнюк В. В. Средства трансдисциплінарного представлення інформаційних ресурсів різних стилей. *Information Technologies and Knowledge*. 2020. Vol. 9, No 1. P. 78-99.
5. **Ресурс онтологій:** <https://covid19tdm.ulif.org.ua/>  
<https://e-devel.ulif.org.ua/?privategraph=1366&view=graphedit>  
<https://e-devel.ulif.org.ua/?sharedgraph=6344650789599&view=graphedit>  
<https://e-devel.ulif.org.ua/?sharedgraph=634464fd05cfb&view=graphedit>  
<https://e-devel.ulif.org.ua/?sharedgraph=6076e4b350e5e>  
<https://e-devel.ulif.org.ua/?sharedgraph=60d08aec7c58d>
6. **Терміни впровадження:** 2019 – 2022 рр.
7. **Ефективність впровадження:** Підвищення ефективності автоматичної класифікації та відповідного пошуку інформації в природномовних виданнях та розширення ряду наявних засобів представлення мережевих документів в електронному вигляді. Розробка ІТ-технології формування онтологій текстів на основі репрезентації їх таксономій, зв'язків між їх контекстами та множинного представлення текстових джерел, що забезпечує ефективний пошук інформації у природномовних текстах. Створення онтологічного репозиторію нормативних документів.
8. **Зауваження, пропозиції:** Розповсюдження отриманих позитивних результатів впровадження шляхом застосування технологічних засобів онтолого-керованого трансдисциплінарного представлення великих масивів слабо- і неструктурованої інформації для оптимізації роботи експертів за необхідності вилучення із множини текстових документів знань, релевантних до заданої предметної галузі, здійснення їх системно-онтологічної структуризації, а також прийняття рішень.

Відповідальний за впровадження:  
 Завідувач відділу інформатики



М.В. Надутенко



ЗАТВЕРДЖУЮ  
 Перший проректор НУОЗ України  
 професор І. Л. Шурика

Ю. П. Вдовиченко

**АКТ**  
**впровадження науково-технічної продукції**

**1. Найменування пропозиції для впровадження:** Науково-технологічні та методичні засоби онтолого-керованого трандисциплінарного представлення великих масивів слабко і неструктурованої інформації, щодо підтримки процесів пошуку, класифікації та вибору необхідного контенту. Створення знання-орієнтованих персональних бібліотек та репозиторіїв. Виявлення критеріальних ознак щодо оцінювання та прийняття рішень на основі методології системно-онтологічного аналізу інформаційних ресурсів.

**2. Ким запропоновано, адреса:** відділ інформаційної безпеки Інституту телекомунікацій і глобального інформаційного простору НАН України; адреса: Чоколівський б-р, 13, 03186, м. Київ.

**3. Виконавець:** ГАЙКО Світлана Іванівна.

**4. Джерело інформації:**

Стрижак, О.С., Приходнюк В.В., Гайко С. І., Шаповалов В.Б. Відображення мережевої інформації у вигляді інтерактивних документів. Трандисциплінарний підхід. *Математичні моделі в економіці*. 2018. № 3 (12). С. 87-100.

Приходнюк В.В., Стрижак О.С., Гайко С. І., Чепков Р.І. Інформаційно-аналітичний комплекс підтримки процесів трандисциплінарних досліджень. *Екологічна безпека та природокористування*. 2018. № 4 (28). С. 103-119.

Гайко С. І., Приходнюк В. В. Автоматизована структуризація освітніх ресурсів. *Наукові записки МАНУ*. 2021. № 1.

Гайко С. І., Приходнюк В. В. Средства трандисциплінарного представлення інформаційних ресурсів різних стилей. *Information Technologies and Knowledge*. 2020. Vol. 9, No 1. P. 78-99.

**5. Реєстр онтологій:** <https://covid19tdm.ulif.org.ua/>

<https://e-devel.ulif.org.ua/?privategraph=1366&view=graphedit>

<https://e-devel.ulif.org.ua/?sharedgraph=6344650789599&view=graphedit>

<https://e-devel.ulif.org.ua/?sharedgraph=634464fd05efb&view=graphedit>

<https://e-devel.ulif.org.ua/?sharedgraph=6076e4b350e5e>

<https://e-devel.ulif.org.ua/?sharedgraph=60d08acc7c58d>

**6. Терміни впровадження:** 2019 – 2023 рр.

**7. Ефективність впровадження:** Підвищення ефективності автоматичної класифікації та відповідного пошуку інформації в природномовних виданнях та розширення ряду наявних засобів представлення мережевих документів в електронному вигляді. Розробка ІТ-технології формування онтологій видань на основі репрезентації їх таксономій, зв'язків між їх контекстами та множинного представлення видань, що забезпечує ефективний пошук інформації у природномовних виданнях. Створення онтологічного репозиторію нормативних документів.

**8. Зауваження, пропозиції:** Розповсюдження отриманих позитивних результатів впровадження шляхом застосування технологічних засобів онтолого-керованого трандисциплінарного представлення великих масивів слабко і неструктурованої інформації для оптимізації роботи експертів при необхідності вилучення із множини текстових документів знань, релевантних до заданої предметної галузі, здійснення їх системно-онтологічної структуризації, а також прийняття рішень.

Завідувач кафедри інформатики,  
 інформаційних технологій та  
 трандисциплінарного навчання  
 професор

О. П. Мінцер

« \_\_\_\_ » \_\_\_\_\_ 20\_\_ р



ЗАТВЕРДЖУЮ  
В.в. директора

Тетяна ПЕЩЕРІНА

АКТ

**впровадження науково-технічної продукції**

1. **Найменування пропозиції для впровадження:** науково-технологічні та методичні засоби онтолого-керованого трансдисциплінарного представлення великих масивів слабо і неструктурованої інформації щодо підтримки процесів пошуку, класифікації та вибору необхідного контенту. Створення знання-орієнтованих персональних бібліотек та репозиторіїв. Виявлення критеріальних ознак щодо оцінювання та прийняття рішень на основі методології системно-онтологічного аналізу інформаційних ресурсів.
2. **Ким запропоновано, адреса:** відділ інформаційної безпеки Інституту телекомунікацій і глобального інформаційного простору НАН України; адреса: Чоколівський б-р, 13, 03186, м. Київ.
3. **Виконавець:** ГАЙКО Світлана Іванівна.
4. **Джерело інформації:**
  - Стрижак, О.С., Приходнюк В.В., Гайко С. І., Шаповалов В.Б. Відображення мережевої інформації у вигляді інтерактивних документів. Трансдисциплінарний підхід. *Математичні моделі в економіці*. 2018. № 3 (12). С. 87-100;
  - Приходнюк В.В., Стрижак О.С., Гайко С. І., Чепков Р.І. Інформаційно-аналітичний комплекс підтримки процесів трансдисциплінарних досліджень. *Екологічна безпека та природокористування*. 2018. № 4 (28). С. 103-119;
  - Гайко С. І., Приходнюк В. В. Автоматизована структуризація освітніх ресурсів. *Наукові записки МАНУ*. 2021. № 1;
  - Гайко С. И., Приходнюк В. В. Средства трансдисциплинарного представлення інформаційних ресурсів різних стилей. *Information Technologies and Knowledge*. 2020. Vol. 9, No 1. P. 78-99.
5. **Ресурс онтологій:** <https://covid19tdm.ulif.org.ua/>  
<https://e-devel.ulif.org.ua/?privategraph=1366&view=graphedit>  
<https://e-devel.ulif.org.ua/?sharedgraph=6344650789599&view=graphedit>  
<https://e-devel.ulif.org.ua/?sharedgraph=634464fd05cfb&view=graphedit>  
<https://e-devel.ulif.org.ua/?sharedgraph=6076e4b350e5e>  
<https://e-devel.ulif.org.ua/?sharedgraph=60d08aec7c58d>
6. **Терміни впровадження:** 2019-2022 рр.
7. **Ефективність впровадження:** підвищення ефективності автоматичної класифікації та відповідного пошуку інформації в природномовних виданнях та розширення ряду наявних засобів представлення мережевих документів в електронному вигляді. Розробка ІТ-технології формування онтологій видань на основі репрезентації їх таксономій, зв'язків між їх контекстами та множинного представлення видань, що забезпечує ефективний пошук інформації у природномовних виданнях. Створення онтологічного репозиторію нормативних документів.
8. **Зауваження, пропозиції:** розповсюдження отриманих позитивних результатів впровадження шляхом застосування технологічних засобів онтолого-керованого трансдисциплінарного представлення великих масивів слабо і неструктурованої інформації для оптимізації роботи експертів при необхідності вилучення із множини текстових документів знань, релевантних до заданої предметної галузі, здійснення їх системно-онтологічної структуризації, а також прийняття рішень.

Відповідальний за впровадження,  
завідувач відділу створення і використання  
інтелектуальних мережних інструментів, к.т.н.  
\_\_\_\_\_ 2023

*В.В.*

Віталій ПРИХОДНЮК

## Додаток Б. Список опублікованих праць за темою дисертації

### Публікації у наукових фахових виданнях України

1. Стрижак, О.Є., Приходнюк В.В., **Гайко С. І.**, Шаповалов В.Б. Відображення мережевої інформації у вигляді інтерактивних документів. Трансдисциплінарний підхід. *Математичне моделювання в економіці*. 2018. № 3 (12). С. 87-100.  
ISSN 2409-8876  
[https://itgip.org/wp-content/uploads/2019/02/%D1%80%D0%B5%D0%B4-%D0%9C%D0%9C%D0%95\\_3\\_2018.pdf](https://itgip.org/wp-content/uploads/2019/02/%D1%80%D0%B5%D0%B4-%D0%9C%D0%9C%D0%95_3_2018.pdf)
2. Приходнюк В.В., Стрижак О.Є., **Гайко С. І.**, Чепков Р.І. Інформаційно-аналітичний комплекс підтримки процесів трансдисциплінарних досліджень. *Екологічна безпека та природокористування*. 2018. № 4 (28). С. 103-119.  
ISSN 2411-4049  
<https://itgip.org/wp-content/uploads/2021/03/2018-28.pdf>
3. **Гайко С. І.**, Приходнюк В. В. Підхід до автоматизованої структуризації освітніх ресурсів на основі методу рекурсивної редукції. *Наукові записки Малої академії наук України*. 2021. № 1 (19). С. 28-38. DOI: <http://doi.org/10.51707/2618-0529-2021-20-03>.  
ISSN 2786-4510 (Online)  
<http://snman.science/index.php/sn/issue/view/8/8>
4. **Гайко С. І.** Онтолого-керовані засоби оброблення та подання великих масивів неструктурованих текстів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 2 (24). С. 27-38. DOI: <https://doi.org/10.30837/ITSSI.2023.24.027>.  
ISSN 2524-2296 (Online)  
<https://itssi-journal.com/index.php/itssi/article/view/387>

### Публікації у виданнях іноземних держав, що входять до організації економічного співробітництва та розвитку Європейського союзу

5. **Гайко С. І.**, Приходнюк В. В. Средства трансдисциплінарного представлення інформаційних ресурсів різних стилей. *Information Models and Analyses*. 2020. Vol. 9, No 1. P. 78-99.  
ISSN 1314-6432 (Online)  
<http://www.foibg.com/ijima/vol09/ijima09-01-p05.pdf>

*Статті у наукових закордонних виданнях, що індексуються у міжнародних наукометричних базах (Scopus, Web of Science)*

6. Stryzhak O., Prychodniuk V., Popova M., Nadutenko M., **Haiko S.**, Shepkov R. (2021) Development of an Oceanographic Databank Based on Ontological Interactive Documents. In: Arai K. (eds) *Intelligent Computing. Lecture Notes in Networks and Systems*. Vol. 284. Springer, Cham. P. 97-114. DOI: [https://doi.org/10.1007/978-3-030-80126-7\\_8](https://doi.org/10.1007/978-3-030-80126-7_8).  
ISSN 2367-3389 (Online)  
[https://link.springer.com/chapter/10.1007/978-3-030-80126-7\\_8](https://link.springer.com/chapter/10.1007/978-3-030-80126-7_8)

*Матеріали та тези конференцій*

7. **Гайко С. І.**, Приходнюк В. В. Автоматизована обробка академічних видань засобами системи «Рекурсивний редуктор». *Інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях*: колективна монографія за мат. XVII міжнар. наук.-практ. конф. (м. Київ, 25-26 вересня 2018). Київ, 2018. С. 126-128.  
<https://itgip.org/wp-content/uploads/2018/10/%D0%97%D0%B1%D1%96%D1%80%D0%BA%D0%B0-%D0%BC%D0%B0%D0%BA%D0%B5%D1%82-2018-1.pdf>
8. Гуржій А., **Гайко С.**, Попова М., Стрижак О. Трансдисциплінарні когнітивні засоби підтримки наукових досліджень. *Освіта для миру = Edukacja dla pokoju*: зб. наук. пр. VIII українсько-польського наук. форуму (м. Київ, 8-10 жовтня 2019 р.). Київ, 2019. Т.1. С. 190-200.  
[http://ipood.com.ua/data/UAPLForum2019/volume\\_1\\_2019.pdf](http://ipood.com.ua/data/UAPLForum2019/volume_1_2019.pdf)
9. **Гайко С. І.**, Приходнюк В. В. Обробка інформації в середовищі «ТОДОС-Редуктор» на прикладі тексту публіцистичного стилю. *Радіоелектроніка та молодь у XXI столітті*: матеріали 23-го міжнар. молодіжного форуму (м. Харків, 16-18 квітня 2019 р.). Харків, 2019. С. 11-12.  
<https://nure.ua/wp-content/uploads/workshop/konferentsiia-upravlinnia-znanniamyta-konkurentna-rozvidka-.pdf>
10. **Гайко С. І.** Методика застосування системи «Рекурсивний редуктор». *Сучасні технології в науці та освіті*: матеріали другої міжнар. науково-практ. конф. (м. Сєверодонецьк, 5–7 березня 2019 р.). Сєверодонецьк, 2019. С. 90-92.  
[dSPACE.snu.edu.ua](https://dSPACE.snu.edu.ua)

11. **Гайко С. І.,** Приходнюк В. В. Редуктивні правила обробки відомчої інформації. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях:* колективна монографія за матеріалами XVIII міжнар. наук.-практ. конф. (м. Київ, 01-02 жовтня 2019 р.). Київ, 2019. С. 196-199.  
[https://itgip.org/wp-content/uploads/2019/10/1\\_%D0%9A%D0%BD%D0%B8%D0%B3%D0%B0\\_%D1%81%D0%B0%D0%B9%D1%82.pdf](https://itgip.org/wp-content/uploads/2019/10/1_%D0%9A%D0%BD%D0%B8%D0%B3%D0%B0_%D1%81%D0%B0%D0%B9%D1%82.pdf)
12. **Гайко С. І.,** Приходнюк В. В. Застосування методу рекурсивної редукції для обробки великих масивів різнорідних документів. *Радіоелектроніка та молодь у XXI столітті:* матеріали 24-го Міжнар. молодіжного форуму (м. Харків, 7-9 квітня 2020 р.). Харків, 2020. С. 181-182.  
[https://drive.google.com/file/d/1EwFV6hpQQSHa2xe\\_VfpTeQevmqlntVRw/view](https://drive.google.com/file/d/1EwFV6hpQQSHa2xe_VfpTeQevmqlntVRw/view)
13. **Гайко С. І.,** Приходнюк В. В. Представлення освітніх ресурсів у вигляді інтерактивних документів. *Інтернет – Освіта – Наука – 2020:* зб. праць XII міжнар. наук.-практ. конф. (м. Вінниця, 26 – 29 травня 2020 р.). Вінниця, 2020. С. 249-251.  
<https://ies.vntu.edu.ua/reports/program/WORK-IES-2020.pdf>
14. **Гайко С. І.,** Приходнюк В. В. Засоби трансдисциплінарного представлення освітніх ресурсів. *Актуальні аспекти розвитку STEM-освіти у навчанні природничо-наукових дисциплін:* зб. мат. III міжнар. наук.-практ. конф. (м. Кропивницький, 14-15 травня 2020 р.). Кропивницький, 2020. С. 61-64.  
<https://www.glau.kr.ua/index.php/ua/home-ua/2247-1415marchua>
15. **Гайко С. І.,** Зотова Л. В. Структуризація інформаційних ресурсів різних стилів у форматі наративного дискурсу. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: тенденції 2020 року:* колективна монографія за матеріалами XIX міжнар. наук.-практ. конф. (м. Київ, 06-07 жовтня 2020 р.). Київ, 2020. С. 175-177.  
[https://itgip.org/wp-content/uploads/2020/10/%D0%97%D0%B1%D1%96%D1%80%D0%BA%D0%B0\\_2020\\_1.pdf](https://itgip.org/wp-content/uploads/2020/10/%D0%97%D0%B1%D1%96%D1%80%D0%BA%D0%B0_2020_1.pdf)

16. **Гайко С. І.**, Стрижак О. Є. Модель трансдисциплінарної інтеграції інформаційних ресурсів різних стилів. *Ідеї академіка В.М. Глушкова і сучасні проблеми теоретичної кібернетики*: матеріали ІХ всеукр. наук.-практ. конф. (м. Київ, 18 грудня 2020 р.). Київ, 2020. С. 23-26.

[https://tc.csc.knu.ua/wp/wp-content/uploads/2021/03/%D0%97%D0%B1%D1%96%D1%80%D0%BD%D0%B8%D0%BA-%D0%93%D0%A7-2020\\_23.03.2021.pdf](https://tc.csc.knu.ua/wp/wp-content/uploads/2021/03/%D0%97%D0%B1%D1%96%D1%80%D0%BD%D0%B8%D0%BA-%D0%93%D0%A7-2020_23.03.2021.pdf)

17. **Гайко С. І.** Застосування системи автоматизованої обробки різностильових документів на прикладі океанографічних законодавчих актів. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: тенденції 2020 року*: колективна монографія за матеріалами ХІХ міжнар. наук.-практ. конф. (м. Київ, 04-07 жовтня 2020 р.). Київ, 2021. С. 203-206.

[https://itgip.org/wp-content/uploads/2021/10/1\\_%D0%97%D0%B1%D1%96%D1%80%D0%BA%D0%B0\\_2021.pdf](https://itgip.org/wp-content/uploads/2021/10/1_%D0%97%D0%B1%D1%96%D1%80%D0%BA%D0%B0_2021.pdf)

*Авторське свідоцтво*

18. А. с. Комп'ютерна програма «Трансдисциплінарні онтологічні дослідження операціональних середовищ і процесів» («ТОДОС-Процеси») / О. Є. Стрижак, В. Ю. Величко, В. В. Приходнюк, М. В. Надутенко, В. В. Горборуков, О. В. Франчук, С. О. Довгий, О. В. Лісовий, О. В. Палагін, І. В. Сергієнко, В. А. Широков, Ю. О. Романенков, І. В. Шостак, І. Б. Чепков, А. Ю. Гупало, Г. М. Потапов, Д. А. Філістєєв, Л. С. Глоба, О. В. Гладун, Т. О. Грязнухіна, М. Ю. Кригін, Т. П. Любченко, І. В. Шевченко, Л. А. Шипіцина, **С. І. Гайко**, К. С. Малахов (Україна). - № 96130; Дата реєстрації 18.02.2020.

<https://its.kpi.ua/uk/patenty%20NDI%20TK>



### Додаток В. Відомості про апробацію результатів дисертації

№ п/п	Назви конференції, конгресу, симпозіуму, семінару, школи	Місце проведення	Дата проведення	Форма участі
1	Інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях	Київ	25-26 вересня 2018 р.	очна
2	Освіта для миру = <i>Edukacja dla pokoju</i>	Київ	8-10 жовтня 2019 р.	заочна
3	Радіоелектроніка та молодь у XXI столітті	Харків	16-18 квітня 2019 р.	заочна
4	Сучасні технології в науці та освіті	Сєверодонецьк	5-7 березня 2019 р.	очна
5	Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях	Київ	01-02 жовтня 2019 р.	заочна
6	Радіоелектроніка та молодь у XXI столітті	Харків	7-9 квітня 2020 р.	заочна
7	Інтернет –Освіта – Наука – 2020	Вінниця	26-29 травня 2020 р.	очна
8	Актуальні аспекти розвитку STEM-освіти у навчанні природничо-наукових дисциплін:	Кропивницький	14-15 травня 2020 р.	очна
9	Ідеї академіка В.М. Глушкова і сучасні проблеми теоретичної кібернетики	Київ	18 грудня 2020 р.	очна
10	Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: тенденції 2020 року	Київ	04-07 жовтня 2020 р.	очна
11	Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: виклики 2021 року	Київ	04-08 жовтня 2021 р.	заочна

## Додаток Г. Експертна оцінка підстилів

Критерії Підстилі	Сутності	Відношення	Атрибути	Контексти	Зміст, рубрикація	Роздільники	Списки	Спеціалізовані об'єкти
<i>Власне науковий</i>	висока складність	висока складність	висока складність	висока складність	допустима складність	висока складність	допустима складність	допустима складність
<i>Науково-довідковий</i>	висока якість	висока якість	допустима складність	висока якість	висока якість	висока якість	висока якість	допустима складність
<i>Науково-навчальний</i>	допустима складність	допустима складність	допустима складність	допустима складність	висока якість	висока складність	допустима складність	допустима складність
<i>Науково-популярний</i>	висока складність	низька наявність	висока складність	висока складність	допустима складність	низька важливість	низька наявність	низька важливість
<i>Законодавчий</i>	допустима складність	допустима складність	допустима складність	висока якість	висока якість	висока якість	допустима складність	низька наявність
<i>Дипломатичний</i>	допустима складність	низька наявність	висока складність	висока складність	допустима складність	низька важливість	висока складність	низька наявність
<i>Адміністративно-канцелярський</i>	висока складність	допустима складність	висока складність	допустима складність	висока складність	висока складність	допустима складність	низька наявність
<i>Юридичний</i>	допустима складність	висока складність	низька наявність	висока складність	висока складність	низька важливість	низька наявність	низька наявність
<i>Власне публіцистичний</i>	висока складність	низька наявність	низька наявність	висока складність	висока складність	низька важливість	низька наявність	низька наявність
<i>Науково-публіцистичний</i>	висока складність	низька наявність	висока складність	висока складність	висока складність	низька важливість	низька наявність	низька наявність
<i>Художньо-публіцистичний</i>	висока складність	низька наявність	низька наявність	висока складність	низька наявність	низька важливість	низька наявність	низька наявність
<i>Мова преси</i>	висока складність	низька наявність	низька наявність	висока складність	висока складність	низька важливість	низька наявність	низька наявність
<i>Мова радіо і телебачення</i>	висока складність	низька наявність	низька наявність	висока складність	низька наявність	низька важливість	низька наявність	низька наявність
<i>Усна публічна мова</i>	висока складність	низька наявність	низька наявність	висока складність	низька наявність	низька наявність	низька наявність	низька наявність
<i>Церемоніальна мова</i>	висока складність	низька наявність	допустима складність	висока складність	низька наявність	низька наявність	низька наявність	низька наявність
<i>Епічний, прозовий</i>	висока складність	низька наявність	низька важливість	висока складність	низька важливість	низька важливість	низька наявність	низька наявність

