

В. В. Приходнюк,
О. С. Кузьменко,
В. В. Горборуков

ОНТОЛОГОКЕРОВАНА ОБРОБКА СТАТИСТИЧНИХ ДАНИХ ЗІ ЗВІТІВ РЕГІОНАЛЬНИХ ВІДДІЛЕНЬ НЦ «МАНУ»

Анотація. У статті представлено результати дослідження, спрямованого на обґрунтування концептуальних і методологічних основ створення онтологокерованої системи обробки статистичних даних зі звітів регіональних відділень Національного центру «Мала академія наук України» (далі — НЦ «МАНУ»). Актуальність дослідження зумовлена потребою автоматизації збору та обробки великих обсягів освітньої інформації, що сприятиме підвищенню прозорості, достовірності та аналітичної цінності звітних матеріалів для управлінських рішень. На основі аналізу сучасних підходів до обробки статистичних звітів у діяльності НЦ «МАНУ» виокремлено ключові проблеми: відсутність уніфікованих структур даних, фрагментарність інформації та складність її порівняльного аналізу. У роботі запропоновано концепцію онтологокерованої системи, яка забезпечує формалізований опис предметної області, автоматизований збір, семантичну обробку, трансформацію та верифікацію статистичних даних. Обґрунтовано методологічні засади побудови такої системи, зокрема принципи онтологічного моделювання, інтероперабельності, семантичної узгодженості та інтеграції з іншими цифровими сервісами. Розглянуто технологічні рішення для реалізації системи, а також оцінено її ефективність у підвищенні якості аналітики та стратегічного планування в межах НЦ «МАНУ». Окреслено можливості масштабування запропонованої моделі на інші освітні та наукові установи України шляхом адаптації онтології до відповідних предметних галузей. Результати дослідження мають практичну цінність для формування інноваційної освітньої аналітики та цифрового управління в освітньо-науковому середовищі.

Ключові слова: онтологокеровані системи, освітня статистика, Мала академія наук України, цифрова трансформація, семантична інтеграція, освітня аналітика, стратегічне планування.

Постановка проблеми. У сучасних умовах розвитку інформаційного суспільства особливо значення набувають формалізація, структуризація та інтелектуальний аналіз статистичних даних, що накопичуються в результаті функціонування освітніх та науково-дослідницьких установ. Однією з таких ключових інституцій є Національний центр «Мала академія наук України» (далі — НЦ «МАНУ») — наукова установа, яка координує діяльність регіональних

відділень і забезпечує підтримку наукового розвитку учнівської молоді. У процесі вивчення діяльності регіональних осередків НЦ «МАНУ» щорічно збираються звіти, що відображають результати участі учнівської молоді у науково-дослідницьких проєктах, конкурсах, олімпіадах, конференціях та інших заходах. Такі звіти містять як кількісні, так і якісні показники, що є важливою основою для стратегічного планування, оцінювання ефективності освітніх програм, а також виявлення регіональних потреб і пріоритетів. Однак сучасні підходи до збору, обробки та аналізу цієї інформації не завжди відповідають

вимогам цифрової епохи та потенціалу новітніх інформаційно-аналітичних технологій.

Отже, виокремимо основні виклики, з якими стикається система статистичного звітування та аналізу у сфері діяльності НЦ «МАНУ», а саме:

- *відсутність стандартизованості структури звітів* — формати звітів, що надходять із різних регіонів, часто різняться за структурою, змістом та рівнем деталізації, а це ускладнює централізований аналіз, інтеграцію даних та порівняння їх між собою;
- *фрагментарність та неоднорідність даних* — дані, що подаються у вигляді таблиць, графіків, текстових описів, часто не мають єдиної логічної моделі, що призводить до втрати важливої інформації під час обробки й аналізу;
- *низький рівень автоматизації аналітичних процедур* — здебільшого аналітика звітів здійснюється ручним способом або за допомогою базових статистичних інструментів, що не дає змоги глибоко аналізувати тренди, виявляти закономірності або прогнозувати майбутні показники;
- *високий поріг входження до аналізу* — відсутність єдиного структурованого середовища для роботи з даними значно ускладнює підготовку аналітиків нового покоління, які мали б здійснювати якісний аналіз на основі багатовимірних звітних даних;
- *відсутність семантичної інтеграції даних* — показники, які мають однакові назви, можуть мати різне значення в контексті різних регіонів або років, що створює бар'єр для уніфікованого аналізу;
- *недостатня візуалізація результатів* — звіти не завжди супроводжуються інтерактивними або динамічними засобами візуалізації, що ускладнює розуміння їх змісту для широкого кола стейкхолдерів.

У контексті означених викликів перспективним напрямом розв'язання проблеми є впровадження онтологокерованого підходу до обробки статистичних звітів регіональних відділень НЦ «МАНУ». Такий підхід передбачає розроблення і застосування онтологічної моделі предметної області, яка б слугувала семантичним каркасом для уніфікації та формалізації даних, а також забезпечувала б можливості для автоматизованого аналізу, класифікації та виявлення нових знань.

Зазначимо, що онтології являють собою формалізовані моделі знань, які описують об'єкти

предметної області, їх атрибути, відношення та правила взаємодії. Тому застосування онтологій у процесі обробки статистичних звітів дає змогу: забезпечити єдине семантичне представлення даних з усіх регіонів; автоматизувати процеси збору, верифікації та інтеграції інформації; уніфікувати термінологію і структуру показників; забезпечити інтероперабельність з іншими освітніми системами та базами даних; підвищити якість і точність аналітичних висновків; створити умови для побудови адаптивних інтелектуальних сервісів, що підтримують прийняття рішень на основі даних.

Отже, постає проблема розроблення системи онтологокерованої обробки статистичних даних, яка б поєднувала інструменти штучного інтелекту, аналізу даних, семантичного моделювання та візуалізації інформації. Така онтологокерована система має охоплювати: створення онтології предметної області діяльності НЦ «МАНУ», розроблення моделей трансформації неструктурованих звітів у формалізовані дані, впровадження інструментів семантичного аналізу та логічного виводу, створення інтерфейсів користувача для візуалізації результатів.

Мета статті — обґрунтувати концептуальні та методологічні засади створення онтологокерованої системи обробки статистичних даних зі звітів регіональних відділень НЦ «МАНУ», яка забезпечить уніфікацію, формалізацію, автоматизований аналіз та візуалізацію інформації для підвищення ефективності управління та стратегічного планування.

Основні **завдання** дослідження: проаналізувати наявні підходи до збору, структурування та обробки статистичних звітів у регіональних відділеннях НЦ «МАНУ» з позиції їх готовності до інтеграції в онтологічну систему; обґрунтувати концептуальні та методологічні основи створення онтологокерованої системи обробки статистичних даних зі звітів регіональних відділень НЦ «МАНУ»; розробити й обґрунтувати онтологокеровану систему, яка забезпечує автоматизований збір, трансформацію, семантичну інтеграцію та верифікацію статистичних даних; оцінити перспективи впровадження онтологокерованої системи в практику діяльності НЦ «МАНУ» і визначити можливості її масштабування на інші освітні та наукові установи.

Аналіз наукових досліджень. Вивчення сучасних підходів до збору, структурування

та обробки статистичних звітів у регіональних відділеннях НЦ «МАНУ» з позиції їх готовності до інтеграції в онтологічну систему вимагає розгляду актуальних методологій, інструментів та технологій, що застосовуються в суміжних сферах освіти, управління даними та штучному інтелекті.

Найперше у працях А. Ганджімі, В. Пресутті [1] наголошується на важливості чітких принципів онтологічного моделювання для ефективної семантичної інтеграції розрізнених даних. Застосування формалізованих онтологій дає змогу уніфікувати термінологію та структуру звітів, що особливо актуально у випадку різномірних регіональних форматів звітності. Дослідження З. Лю, Х. Лі, Ю. Вана [2] демонструють, що автоматизація вилучення та анотації статистичних даних із текстових і табличних звітів за допомогою методів машинного навчання й обробки природної мови (NLP) значно підвищує ефективність підготовки даних для подальшого аналізу в онтологічних системах. Значна увага в сучасних дослідженнях приділяється проблемі інтероперабельності даних. Стаття Дж. Чжана, Ю. Хуана, Л. Чена [3] висвітлює концептуальні рамки і технічні рішення для забезпечення сумісності між різними інформаційними системами освіти на основі онтологічних моделей. Це створює умови для централізованого та масштабованого аналізу регіональних звітів. Крім того, в дослідженні Т. Нгуєна, С. Лі, Х. Кіма [4] розглядаються методи візуалізації результатів аналізу статистичних даних із застосуванням онтологій, що сприяє кращому сприйняттю інформації користувачами, зокрема освітніми менеджерами та аналітиками. Останні огляди Б. Сміта, М. Джонсона, П. Гарсії [5] підтверджують, що інтеграція семантичних технологій у системи освіти та моніторингу якості навчання відкриває нові перспективи для створення інтелектуальних сервісів, здатних підтримувати прийняття обґрунтованих управлінських рішень.

Щодо українських наукових досліджень, то розвідки О. Стрижака [6–8] демонструють практичні підходи до розроблення онтологічних моделей для освітніх даних, також у них наголошується на важливості уніфікації та формалізації звітної інформації для покращення процесів її аналізу і прийняття рішень у системі освіти.

Науковці С. Кан, Дж. Кім, Х. Парк [9] описують архітектуру онтологізованих систем,

що забезпечують гнучкий збір та структурування статистичних даних, адаптованих до специфіки регіональних відділень. Це актуально для НЦ «МАНУ», де інформація має бути узгодженою й інтегрованою. У роботі С. Чена, Л. Вана, Ю. Лю [10] проаналізовано застосування онтологічних моделей для інтеграції великих масивів даних із різномірних джерел, що забезпечує високий рівень семантичної сумісності та автоматизації обробки статистичних звітів. Дослідження А. Шарма, В. Гупта [11] окреслюють роль онтологій у покращенні якості даних через процеси валідації та нормалізації інформації, що є критичним для надійності статистичних звітів регіональних структур НЦ «МАНУ». Провідні учені, як-от М. Парк, С. Лі [12], розглядають виклики масштабування онтологічних систем стосовно освітніх організацій, де важливо забезпечити одночасну інтеграцію великої кількості даних, що надходять із регіональних відділень, при цьому зберігаючи цілісність і достовірність інформації. У контексті використання інтелектуальних сервісів для аналізу статистичних звітів, що розглянуто в роботі Р. Еванса, М. Томаса, Д. Вайта [13], продемонстровано, як поєднання онтологічних технологій із методами штучного інтелекту може підвищити рівень автоматизації і точності аналітики.

Отже, аналіз наукових досліджень свідчить, що наявні підходи охоплюють основні аспекти онтологічного моделювання, автоматизації збору та візуалізації статистичних даних, але потребують адаптації до специфіки діяльності регіональних відділень НЦ «МАНУ». Скажімо, недостатньо уваги приділяється уніфікації структури звітів і розробленню гнучких моделей трансформації даних з урахуванням їх різномірності. Тому подальші дослідження мають бути зосереджені на комплексній методології онтологізованої обробки, що поєднує формалізацію предметної області, автоматизований семантичний аналіз, інтерактивну візуалізацію та підтримку прийняття рішень на основі інтегрованих статистичних даних регіональних відділень НЦ «МАНУ».

Виклад основного матеріалу. Створення онтологізованої системи для обробки статистичних звітів регіональних відділень НЦ «МАНУ» потребує комплексного підходу, який ґрунтується на чітко визначених концептуальних і методологічних засадах. Такі засади

є ключовими для забезпечення якісного збору, інтеграції, обробки та аналізу статистичних даних, що надходять із різних регіонів країни.

Концептуальні основи полягають у тому, що:

- онтологія як основний концептуальний інструмент виступає моделлю формалізованих знань, що відображає предметну область із точки зору ключових термінів, категорій, атрибутів та відношень між ними;
- семантична інтеграція дає змогу зберігати глибокий контекст і смислові зв'язки між елементами даних, що особливо важливо стосовно статистичних звітів, у яких дані часто мають багатовимірний характер і містять складні залежності, наприклад, між показниками різних напрямів діяльності регіональних відділень МАНУ;
- онтологічна модель надає можливість не лише зберігати статичні дані, а й підтримувати логічне виведення нових знань на основі наявних фактів. Такий підхід підвищує аналітичний потенціал системи, забезпечуючи додаткові інсайти для прийняття управлінських рішень на різних рівнях.

Методологія створення онтології полягає в детальному аналізі предметної області. Залучення експертів у галузі освіти, статистики та інформаційних технологій є обов'язковим етапом для виявлення ключових концептів і встановлення між ними логічних зв'язків. Паралельно здійснюється ретельний огляд і систематизація наявних статистичних звітів, що дає змогу виявити типові структури, атрибути, а також проблеми гетерогенності та неповноти даних. Після визначення концептуальної моделі здійснюється побудова формальної онтології, що забезпечує

сумісність із іншими інформаційними системами та платформами. Важливим кроком є розроблення правил та обмежень, які регламентують зв'язки між поняттями і перевіряють коректність даних. Тому обробка статистичних звітів охоплює етапи вилучення даних (data extraction), очищення (data cleansing) і трансформації (data transformation), які мають виконуватися відповідно до визначених онтологією стандартів. Застосування автоматизованих алгоритмів на основі штучного інтелекту і машинного навчання може значно підвищити ефективність цих процесів, мінімізуючи людський фактор.

Для аналізу і запитів до бази даних, структурованої за онтологічним принципом, використовуються семантичні мови запитів, як-от SPARQL, що дає змогу формувати складні запити з урахуванням логічних зв'язків між елементами інформації. Це відкриває нові можливості для гнучкого і глибокого аналізу статистичних показників, порівняння даних між різними регіонами та відстеження тенденцій у динаміці. Крім того, методологія передбачає ітеративний підхід до розроблення системи з постійним залученням користувачів — представників регіональних відділень, аналітиків та управлінців. Це дає змогу отримувати зворотний зв'язок, уточнювати функціонал і покращувати користувацький інтерфейс, що є важливим для практичного впровадження системи та підвищення її прийняття серед кінцевих користувачів.

Розглянемо як приклад розроблену нами онтологокеровану систему аналізу структурованої звітної документації НЦ «МАНУ». Об'єкти такої онтології (рис. 1) представляють логічні сутності в рамках масиву даних.

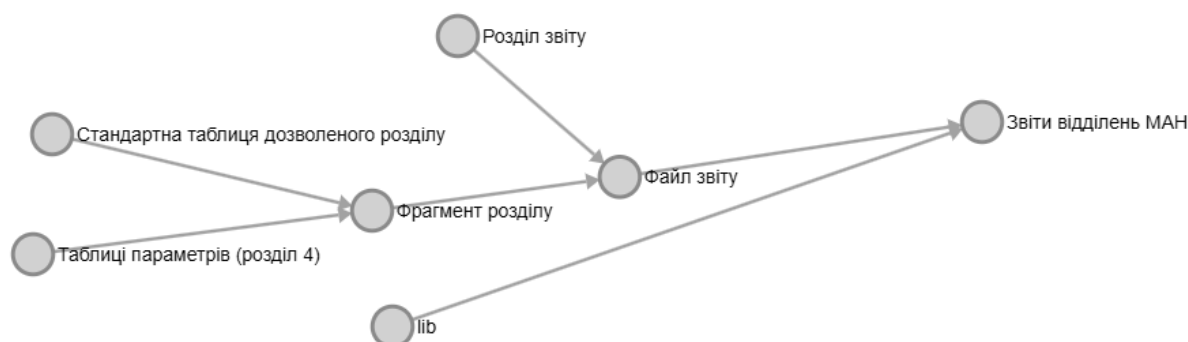


Рис. 1. Структура онтологічного дескриптора

Відповідно до рис. 1. в онтології відображено такі об'єкти:

- «Звіти відділень МАН» — основний вузол, що описує загальну структуру звітів, ініціалізує спеціальні бібліотеки, визначає налаштування для читання DOCX-файлів і загальний алгоритм обробки даних;
- «Файл звіту» — визначає ідентифікацію та спосіб обробки окремих документів, визначає контекст звіту (рік, регіон) і готує текстові розділи для подальшої обробки;
- «Розділ звіту» — відповідає за ідентифікацію заголовків розділів, збереження текстового контенту, визначення року та регіонального відділення на основі попереднього контексту документа;
- «Фрагмент розділу» — обробляє окремі блоки тексту всередині розділів і зберігає структурований контент розділу;
- «Стандартна таблиця дозволеного розділу» — конфігурація для обробки табличних блоків усередині (дозволених) розділів звіту з використанням словників для визначення назв колонок, об'єктів і кількісних значень;
- «Таблиці параметрів (розділ 4)» — спеціалізована конфігурація для обробки таблиць четвертого розділу, де відбувається додаткова попередня трансформація даних для здійснення коректного структурування інформації;
- «lib» — бібліотека допоміжних функцій і тезаурусів, що використовуються для семантичного перетворення та ідентифікації назв розділів, регіонів, колонок таблиць і числових значень.

Об'єкт «Звіти відділень МАН» (рис. 2) є базовим і керує загальною конфігурацією системи для автоматизованого аналізу звітності територіальних відділень МАН:

1. `libraries` — ініціалізація допоміжних бібліотек і модулів, що використовуються для попередньої обробки та семантичного аналізу даних звітів:

- 1) `MANReport()` — основна бібліотека, яка визначає логіку обробки документів типу звітів МАН;
- 2) `XLSX()` — бібліотека для підтримки й обробки файлів формату Excel (XLSX);
- 3) `Ollama()` — бібліотека інтеграції великих мовних моделей (LLM) для обробки природної мови;
- 4) `LLM Embeddings (3, 0.75, embedOllama (bge-m3))` — бібліотека для побудови

семантичних векторних представлень тексту (ембедінгів), що використовує модель Ollama (bge-m3); текст трансформується у вектори із заданим порогом схожості 0.75;

5) `ComplexJSONOutput()` — модуль, що формує складні структуровані результати обробки у форматі JSON.

2. `initialize` — початкова конфігурація для розпізнавання структури та основних елементів звіту. Визначається за допомогою функції `readReportStructure(...)`, яка задає структуру типового звіту МАН і характеризує, як саме варто інтерпретувати його складові:

- 1) повна назва `->name` — визначає назву вузла згідно з типовим звітом;
- 2) `allowTables: addValues` — встановлюються дозволені таблиці розділів типового звіту;
- 3) примітка `->datum`: будь-які додаткові примітки зберігаються як окреме текстове поле.

3. `DOCXReader` — налаштування компонента для читування і розпізнавання вмісту DOCX-документів:

- `DOCXReader(5mm)` — задає параметри читання документів формату DOCX. із зазначенням допустимого форматового параметра (в нашому випадку це відступ у 5 мм), що дає змогу коректно ідентифікувати структурні елементи документів.

4. `PostProcessAllEntities` — визначає механізм післяобробки результатів, отриманих після первинного аналізу звіту:

- `needsAlign(true)` — виконує автоматичне вирівнювання й узгодження всіх отриманих сутностей, щоб забезпечити коректність виводу.

5. `OutputIteration: ComplexOutputIteration(...)` — визначає алгоритм формування результатів для структурованого представлення даних:

- 1) `primaryModel()` — створення основної моделі для результату;
- 2) `modelFromArray(years, Роки звітності, contextLoadArray(years, true, true))` — формування результатів щодо кожного року звітності;
- 3) `ModelFromArray(branches, Територіальні відділення, contextLoadArray (branches, true, true))` — створення окремих моделей результатів стосовно кожного територіального відділення.

6. Об'єкт «lib» (рис. 3) описує допоміжні функції та словники-тезауруси, які використовуються для семантичного аналізу й уточнення даних у процесі обробки звітів територіальних відділень МАН:

- 1) find-node: Fieldset Transformer (...) — виконує семантичну ідентифікацію розділів звіту, визначаючи їх тип і відповідність типовому розділу:
 - а) [Розділ ->text-content]: первинний текст розділу, що виявлено у вхідному документі;
 - б) [Типовий розділ ->chapter, Результат ЛЛМ -> ""]: визначає назву розділу за допомогою результату ЛЛМ або залишає порожнім для подальшого експертного уточнення;
 - в) Semantically Similar Node (...): функція семантичного порівняння з типовими розділами для визначення найбільш ймовірного відповідника;

г) get-region: StringTransformer(Перетворення:: Рєгiони, Назва файлу, Region) — трансформує текстову назву файлу в стандартизовану назву регіону за допомогою відповідного словника.

7. fields — ідентифікація та стандартизація назв колонок таблиць у звіті:

- 1) StringCategorizedTransformer(Перетворення:: Заголовки таблиць, Колонка в таблиці, Значення, Розділ звіту) — функція, що перетворює початкові назви колонок у таблиці на стандартизовані значення за допомогою відповідного словника;
- 2) Names: StringCategorizedTransformer(Перетворення:: Назви, Назва в таблиці, Назва уточнена, Розділ звіту) — уточнює та стандартизує текстові назви, наведені в таблицях, для коректного аналізу в межах певного розділу;

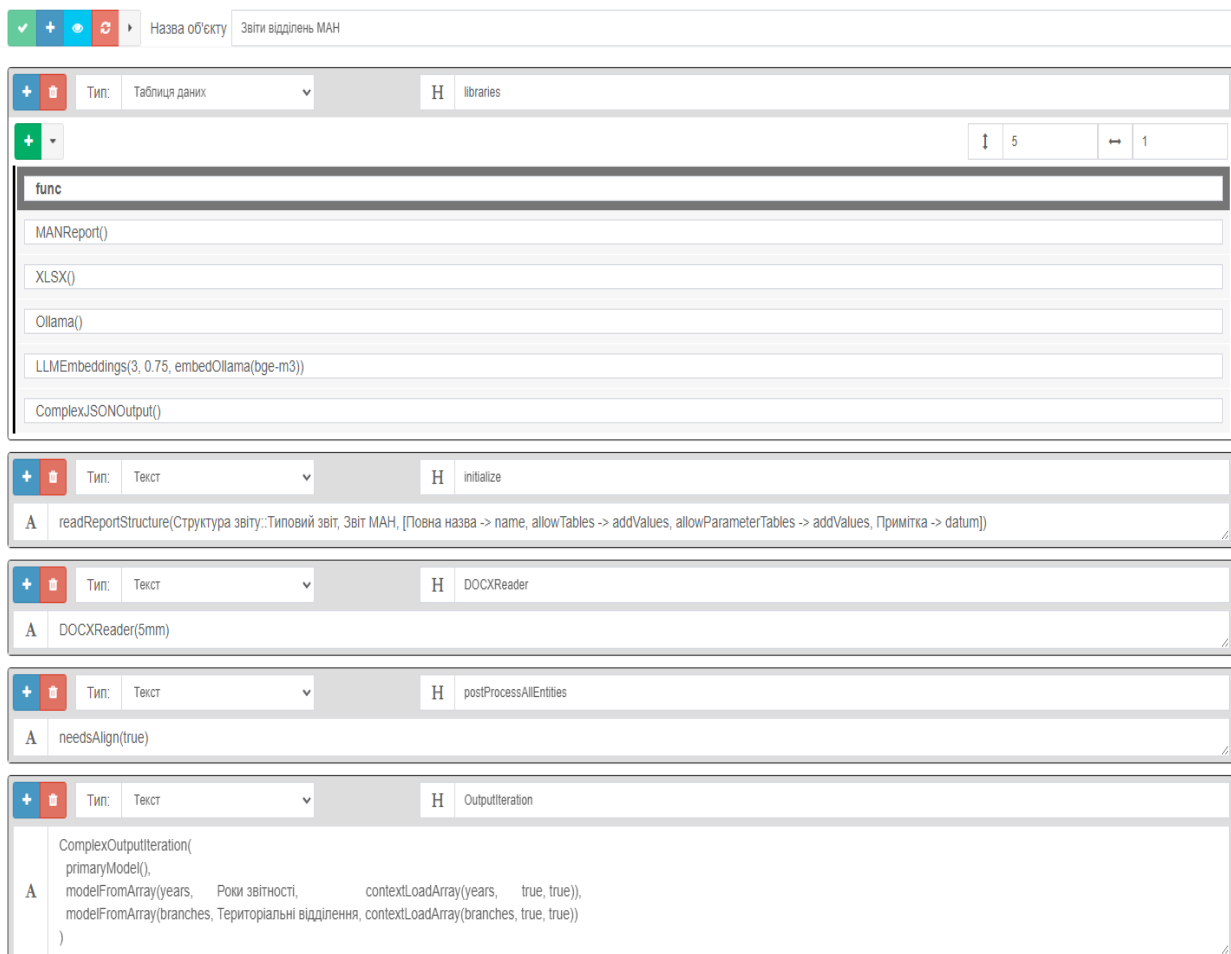


Рис. 2. Атрибути об'єкта «Звіти відділень МАН»

Тип	Значення
Текст	find-node
Текст	get-region
Текст	fields
Текст	names
Текст	counts
Таблиця даних	pre-table-fields

label	func
Відділення	branch
Рік	year

Рис. 3. Атрибути об'єкта «lib»

- 3) counts:String Categorized Transformer (Перетворення:: Кількості, Кількість в таблиці, Кількість уточнена, Поле) — перетворює початкові числові значення в таблицях на стандартизовані й уточнені кількісні показники, тобто здійснює семантичне уточнення кількісних показників у таблицях.
8. pre-table-fields — визначення додаткових спеціальних полів, що додаються до таблиць під час попередньої обробки:
- 1) відділення — стандартизована назва територіального відділення, що визначається за допомогою відповідної функції;
 - 2) рік — визначення року звітності.
9. Об'єкт «Файл звіту» (рис. 4) визначає правила початкової обробки й ідентифікації окремих файлів звітів, забезпечуючи належне структурування та підготовку даних перед подальшим аналізом:

– match Resource: extension (docx) — визначає умови ідентифікації файлів, що підлягають обробці; в цьому випадку встановлюється опрацювання файлів лише у форматі DOCX.

10. preProcessData — етап попередньої обробки даних усередині файлу звіту:

- 1) debug Print Blocks (allFile) — активує режим налагодження, в якому здійснюється виведення вмісту усіх логічних блоків документа для аналізу структури, перевірки коректності розпізнавання та виявлення можливих помилок на етапі обробки;
- 2) Insert Real Headers(...) — процедура автоматичної ідентифікації заголовків розділів, наприклад: validatePattern (text-content, /\d{1,2}\.*/) — виконує перевірку тексту заголовків розділів відповідно

Тип	Значення
Текст	extension(docx)
Група даних	preProcessData
Текст	postProcessData

Рис. 4. Атрибути об'єкта «Файл звіту»

до регулярного виразу («1.», «2.» тощо). А функція `alterFieldset (#lib::find-node)` застосовує семантичне перетворення заголовків за допомогою функції, визначеної в бібліотеці `lib` (`find-node`), для точного визначення назви типового розділу;

- 3) `postProcessData: contextRetain(years, branches)` — зберігає важливий контекстний стан (наприклад, роки звітності й назви територіальних відділень) для подальшого використання та формування остаточного результату.

11. Об'єкт «Розділ звіту» (рис. 5) описує процедуру ідентифікації й обробки окремих розділів усередині звіту, виділяючи їх заголовки і налаштовуючи контекстні параметри для подальшого аналізу даних:

- 1) `match: type(header)` — перевіряє, чи є елемент заголовком розділу; використовується для ідентифікації відповідних блоків у структурі документа;
- 2) `parse: normText(text-content)` — виконує попередню текстову обробку та нормалізацію заголовка розділу (видаляє зайві символи, пробіли тощо), підготує його для подальшого семантичного аналізу.

12. `Transform Fieldset` — визначає перелік додаткових функцій для уточнення інформації в межах поточного розділу:

- 1) `funcField(filename, filename())` — автоматично додає поле з назвою файлу, з якого було взято розділ;
- 2) `funcField(year, rootFoldername(1))` — автоматично додає поле з роком, отриманим із назви теки верхнього рівня, в якій зберігається звіт;
- 3) `thesaurusField([filename -> branch], #lib::get-region)` — застосовує семантичне перетворення назви файлу на стандартизовану назву територіального відділення, використовуючи бібліотеку (`lib::get-region`);
- 4) `contextAddToArray([year -> years, branch -> branches], true)` — долучає отримані параметри року й відділення у відповідний контекстний масив;
- 5) `saveContext([year, branch])` — зберігає поточний контекст для подальшого використання;
- 6) `createContainer: save(false, existNode(text-content))` — визначає створення контейнера для даних розділу (контейнер створюється лише в разі, якщо вузол із поточним текстом заголовка ще не існує);

Назва об'єкту	Розділ звіту
Тип: Текст	match
A	type(header)
Тип: Текст	parse
A	normText(text-content)
Тип: Таблиця даних	transformFieldset
func funcField(filename, filename()) funcField(year, rootFoldername(1)) thesaurusField([filename -> branch], #lib::get-region) contextAddToArray([year -> years, branch -> branches], true) saveContext([year, branch])	
Тип: Текст	createContainer
A	save(false, existNode(text-content))
Тип: Текст	fillContainer
A	contextSaveDatum(text-row, table(Текст розділів, #lib::pre-table-fields, text(Текст, s(""))))

Рис. 5. Атрибути об'єкта «Розділ звіту»

Тип	Назва	Метод
Текст	all()	match
Текст	normText(text-content)	parse
Текст	load(text-row)	createContainer
Текст	appendToCell(Текст, text-content)	fillContainer
Текст	iterateBlockRecursively()	postProcessContainer

Рис. 6. Атрибути об'єкта «Фрагмент розділу»

- 7) fillContainer: contextSaveDatum(...) — зберігає текстову інформацію про розділ у спеціальну вкладену таблицю «Текст розділів». Для здійснення цієї операції також використовуються допоміжні поля, визначені у бібліотеці (lib::pre-table-fields).
13. Об'єкт «Фрагмент розділу» (рис. 6) відповідає за обробку окремих текстових фрагментів усередині розділів звіту, забезпечуючи їх нормалізацію й накопичення для подальшого аналізу та структурування:
- 1) match: all() — дозволяє обробку всіх текстових фрагментів, що містяться в межах поточного розділу;
 - 2) parse: normText(text-content) — виконує нормалізацію тексту (видалення зайвих пробілів, спеціальних символів тощо);
 - 3) createContainer: load(text-row) — завантажує кожен текстовий фрагмент у створений контейнер на рівні окремого рядка;
 - 4) fillContainer: appendToCell(Текст, text-content) — додає нормалізований текстовий вміст у відповідне поле «Текст» контейнера, накопичуючи інформацію з усіх оброблених фрагментів;
 - 5) postProcessContainer: iterateBlockRecursively() — рекурсивно проходить усі вкладені текстові блоки та їх підблоки, забезпечуючи повну обробку та накопичення усього текстового вмісту розділу.
14. Об'єкт «Стандартна таблиця дозволеного розділу» (рис. 7) описує процедуру обробки таблиць (що містяться у дозволених розділах звіту), тобто визначає їх структуру і спосіб очистки та семантичної стандартизації отриманих даних:

Тип	Назва	Метод	I	O
Таблиця даних		match	2	1
Таблиця даних		preProcessBlock	3	1
Текст	table(withoutBlankColumns(categorizedTransformer(#lib::fields, sectionName())))	parse		
Таблиця даних		transformFieldset	6	1
Текст	load()	createContainer		
Текст	table(Дані розділів, #lib::pre-table-fields, #current::table-fields)	fillContainer		
Таблиця даних		table-fields	2	1

Рис. 7. Атрибути об'єкта «Стандартна таблиця дозволеного розділу»

- 1) `match` — визначає умови ідентифікації таблиць, які підлягають обробці:
 - а) `type(table)` — перевіряє, чи є блок даних таблицею;
 - б) `isSection(allowTables)` — дає змогу обробляти таблиці лише з розділів, що відмічені як дозволені (`allowTables`);
 - 2) `preProcessBlock` — попередня підготовка табличних даних перед основною обробкою:
 - а) `removeSheetPrefix()` — видаляє зайві префікси назв, якщо вони присутні;
 - б) `removeFirstRows (!/*навчальний рік.*/ )` — видаляє зайвий початковий рядок таблиці, якщо він містить маркер «навчальний рік»;
 - в) `debugPrintBlock (standardTables)` — використовується для налагодження і дає змогу перевірити, як оброблено блок таблиці на поточному етапі;
 - 3) `parse: table(...)` — визначає процедуру обробки і семантичної структуризації таблиці з використанням таких функцій:
 - а) `withoutBlankColumns(...)` — видаляє порожні стовпці;
 - б) `categorizedTransformer (#lib:::fields, sectionName())` — застосовує семантичне перетворення назв колонок таблиці відповідно до словника полів (`lib:::fields`) з урахуванням контексту розділу;
 - 4) `transformFieldset` — задає додаткові перетворення, які виконуються над витягнутими даними з таблиці:
 - а) `debugPrint(standardTables)` — визначає виведення результатів у режимі відладки для контролю процесу;
 - б) `thesaurusCategorizedField(/Назва.*/, #lib:::names, sectionName())` — задає семантичне перетворення значень колонок із назвами на стандартизовані назви відповідно до словника (`lib:::names`);
 - в) `thesaurusCategorizedField(/Кількість.*/, #lib:::counts)` — визначає перетворення кількісних значень на нормалізовані числові значення за допомогою словника (`lib:::counts`);
 - г) `trackTableContext(/Назва.*/ )` — зберігає контекст таблиці для колонок, що відповідають вказаному шаблону (містять назву), з метою його подальшого використання;
 - д) `transformRequireAnyValue(/(Назва|Кількість).*/ )` — здійснює обов'язкову перевірку наявності хоча б одного значення в колонках «Назва» або «Кількість»;
 - е) `loadContext([year, branch])` — виконує завантаження даних із контексту: року й територіального відділення для таблиці;
 - 5) `createContainer: load()` — виконує створення контейнера для зберігання даних таблиці;
 - 6) `fillContainer: table (Дані розділів, #lib:::pre-table-fields, #current:::table-fields)` — наповнює контейнер створеними структурованими табличними результатами, для чого використовує додаткові поля з бібліотеки (`lib:::pre-table-fields`) та актуальні поля, що визначені для цієї поточної таблиці (`current:::table-fields`);
 - 7) `table-fields` — задає механізм визначення значень та полів таблиці:
 - а) `simpleFieldset(extractTableContextNames(categoryValueArray (#lib:::names, sectionName())))` — отримує стандартизовані назви об'єктів таблиці згідно з відповідним словником;
 - б) `simpleFieldset(categoryValueArray(#lib:::fields, sectionName()))` — отримує перелік стандартизованих назв атрибутів таблиці відповідно до поточного розділу.
15. Об'єкт «Таблиці параметрів (розділ 4)» (рис. 8) визначає конфігурацію для обробки четвертого розділу. Тут реалізується спеціалізований механізм семантичної інтерпретації та структурування даних для такого типу таблиць. Розглянемо ключові моменти:
- 1) `preProcessFieldsetList` — попередня підготовка набору даних таблиці для здійснення подальшої обробки;
 - а) `preTransformFieldsets (#current:::pre-transformers)` — виконує попередню семантичну трансформацію даних за заданими правилами (`pre-transformers`);
 - б) `rowFromTable(/Назва.*/, /Кількість.*/ )` — вибирає лише ті рядки таблиці, які містять поля «Назва» та «Кількість»;

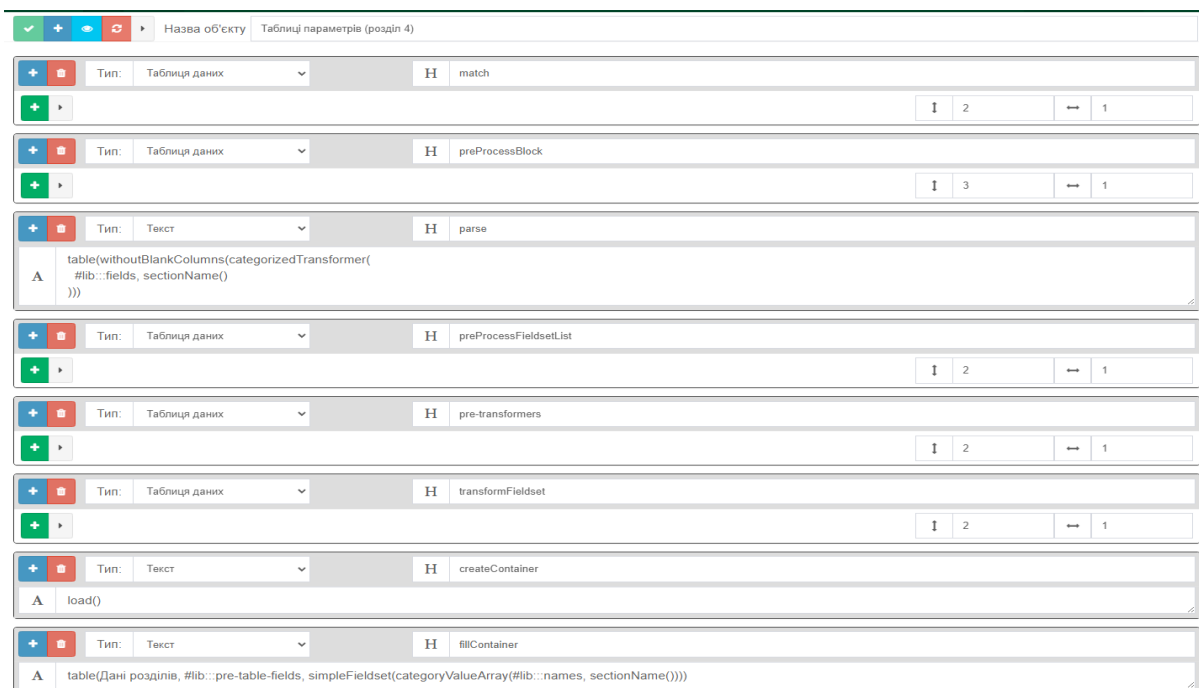


Рис. 8. Атрибути об'єкта «Таблиці параметрів (розділ 4)»

2) pre-transformers — визначає спеціалізовані семантичні трансформації, що застосовуються перед основною обробкою:

- а) thesaurusCategorizedField(/Назва.*/, #lib::names, sectionName()) — семантично уточнює назви рядків відповідно до словника назв (lib::names);
- б) thesaurusCategorizedField(/Кількість.*/, #lib::counts) — стандартизує кількісні значення відповідно до словника кількостей (lib::counts).

У результаті застосування конфігурації формується інформаційна онтологія (рис. 9). Ця онтологія містить таксономічну структуру вузлів розділів типового звіту, в яких уже є цільові дані конкретних МАНівських звітів.

Отже, розроблена нами онтологокерована система дає змогу враховувати внутрішню логіку документів, працює з нестандартними шаблонами, виявляє релевантну інформацію незалежно від розміщення чи назви розділу. Така система надає широкий спектр можливостей, зокрема:

- автоматично структурувати інформацію з документів різного формату (Word, Excel);
- ідентифікувати значущі інформаційні блоки (організаційна структура, кількісні показники, освітні заходи тощо);

- витягати семантичні сутності (напрями, секції, громади, результати та інше);
- відстежувати зв'язки між структурними елементами звітів;
- формувати інтегровані профілі діяльності відділень у часі.

Отже, онтологокерована система забезпечує цілісний, формалізований і масштабований підхід до роботи з великомасштабною звітною інформацією, що критично важливо для аналітичної підтримки прийняття рішень, стратегічного планування та виведення інтегрованих висновків на основі множинних джерел.

Висновки. У контексті цифровізації освітнього та наукового простору України впровадження онтологокерованої системи збору й обробки статистичних даних у практику діяльності НЦ «МАНУ» відкриває нові горизонти як для вдосконалення управлінських процесів, так і для забезпечення якісної аналітики на основі великого обсягу даних, які щорічно акумулюються у звітах регіональних відділень. Практичні переваги впровадження:

1. Створення єдиної онтологічної моделі для всієї мережі НЦ «МАНУ» дасть змогу стандартизувати форми статистичної звітності, подолати проблему семантичної неоднорідності даних і зменшити кількість дублювань та помилок під

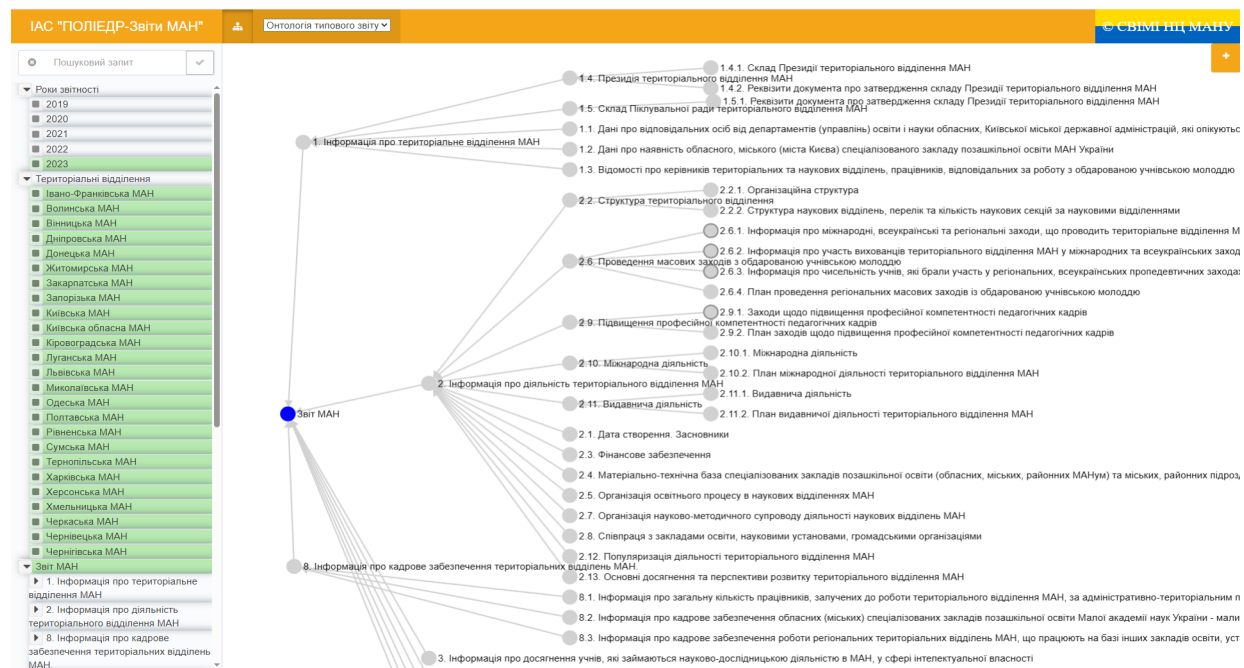


Рис. 9. Інформаційна онтологія, створена в результаті обробки

час внесення інформації ручним способом. За рахунок автоматизованого збору і семантичного аналізу звітних даних можливе формування узагальнених аналітичних моделей, що можуть використовуватися для формування стратегій розвитку освітньо-наукової діяльності.

2. Така система значно підвищує прозорість процесів, уможливлючи відстеження активності учнівської молоді, наукових керівників, напрямів дослідницької роботи в режимі реально часу, оцінювання динаміки результативності участі в конкурсах, проєктах, міжнародній співпраці тощо.

3. Впровадивши онтологічний підхід, можна створити базу знань на основі накопичених даних, яка може бути використана не лише для аналітики, а й для підготовки звітності, внутрішнього моніторингу, зовнішнього аудиту, підготовки грантових заявок, а також для науково-дослідницької діяльності.

У перспективі на основі створеної онтології можна забезпечити інтероперабельність даних між різними платформами і реєстрами (наприклад, «Єдина державна електронна база з питань освіти»), що сприятиме формуванню відкритої екосистеми освітньо-наукових даних в Україні.

Список використаних джерел

- Gangemi A., Presutti V. Ontology Engineering for Semantic Data Integration: Principles and Practice. *Semantic Web Journal*. 2021. Vol. 12. № 3. Pp. 365–387.
- Liu Z., Li H., Wang Y. Automated Semantic Annotation of Statistical Reports Using Machine Learning. *Information Processing & Management*. 2020. Vol. 57. № 6. Article 102341.
- Zhang J., Huang Y., Chen L. Interoperability Frameworks for Educational Data: Ontology-Based Approaches. *Journal of Educational Data Mining*. 2022. Vol. 14. № 2. Pp. 45–68.
- Nguyen T., Lee S., Kim H. Visual Analytics for Educational Data Using Ontologies. *IEEE Transactions on Learning Technologies*. 2019. Vol. 12. № 1. Pp. 15–27.
- Smith B., Johnson M., Garcia P. Semantic Technologies for Education and Training: State-of-the-Art and Challenges. *Computers & Education*. 2023. Vol. 180. Article 104584.
- Стрижак О. Є. Онтологічне моделювання в системах управління знаннями: теорія і практика. *Вісник КНУ імені Тараса Шевченка*. 2021. Вип. 132. С. 57–67.
- Стрижак О. Є. Формалізація освітніх даних у контексті цифрової трансформації. *Науковий вісник НПУ імені М. П. Драгоманова*. 2022. Вип. 45. С. 23–34.
- Стрижак О. Є. Методи уніфікації статистичних звітів на основі онтологічного підходу. *Журнал освітніх технологій*. 2023. Т. 15. № 1. С. 12–26.

9. Kang S., Kim J., Park H. Ontology-driven Data Collection and Structuring System for Regional Educational Offices. *Information Systems*. 2021. Vol. 95. Article 101624.
10. Chen X., Wang L., Liu Y. Semantic Integration of Large-scale Educational Data Using Ontological Models. *Computers in Human Behavior*. 2022. Vol. 131. Article 107243.
11. Sharma A., Gupta V. Enhancing Data Quality in Educational Statistics through Ontology-based Validation. *Journal of Educational Computing Research*. 2020. Vol. 58. № 5. Pp. 960–981.
12. Park M., Lee S. Scalable Ontology-based Framework for Integrating Regional Educational Data. *Expert Systems with Applications*. 2023. Vol. 215. Article 119403.
13. Evans R., Thomas M., White D. AI-Enhanced Ontology Systems for Automated Educational Report Analysis. *Artificial Intelligence in Education*. 2021. Vol. 31. № 3. Pp. 455–472.
6. Stryzhak, O. Ye. (2021). Ontologichne modeliuvannia v systemakh upravlinnia znanniamy: teoriia i praktyka [Ontological modeling in knowledge management systems: theory and practice]. *Visnyk KNU imeni Tarasa Shevchenka — Bulletin of Taras Shevchenko National University of Kyiv*, 132, 57–67 [in Ukrainian].
7. Stryzhak, O. Ye. (2022). Formalizatsiia osvitykh danykh u konteksti tsyfrovoy transformatsii [Formalization of educational data in the context of digital transformation]. *Naukovyi visnyk NPU imeni M. P. Drahomanova — Scientific Bulletin of the Drahomanov National Pedagogical University*, 45, 23–34 [in Ukrainian].
8. Stryzhak, O. Ye. (2023). Metody unifikatsii statystychnykh zvitiv na osnovi ontologichnoho pidkhodu [Methods of unifying statistical reports based on an ontological approach]. *Zhurnal osvitykh tekhnologii — Journal of Educational Technologies*, 15 (1), 12–26 [in Ukrainian].
9. Kang, S., Kim, J., & Park, H. (2021). Ontology-driven Data Collection and Structuring System for Regional Educational Offices. *Information Systems*, 95, 101624.
10. Chen, X., Wang, L., & Liu, Y. (2022). Semantic Integration of Large-scale Educational Data Using Ontological Models. *Computers in Human Behavior*, 131, 107243.
11. Sharma, A., & Gupta, V. (2020). Enhancing Data Quality in Educational Statistics through Ontology-based Validation. *Journal of Educational Computing Research*, 58 (5), 960–981.
12. Park, M., & Lee, S. (2023). Scalable Ontology-based Framework for Integrating Regional Educational Data. *Expert Systems with Applications*, 215, 119403.
13. Evans, R., & Thomas, M., White, D. (2021). AI-Enhanced Ontology Systems for Automated Educational Report Analysis. *Artificial Intelligence in Education*, 31 (3), 455–472.

References

1. Gangemi, A., & Presutti, V. (2021). Ontology Engineering for Semantic Data Integration: Principles and Practice. *Semantic Web Journal*, 12 (3), 365–387.
2. Liu, Z., Li, H., & Wang, Y. (2020). Automated Semantic Annotation of Statistical Reports Using Machine Learning. *Information Processing & Management*, 57 (6), 102341.
3. Zhang, J., Huang, Y., & Chen, L. (2022). Interoperability Frameworks for Educational Data: Ontology-Based Approaches. *Journal of Educational Data Mining*, 14 (2), 45–68.
4. Nguyen, T., Lee, S., & Kim, H. (2019). Visual Analytics for Educational Data Using Ontologies. *IEEE Transactions on Learning Technologies*, 12 (1), 15–27.
5. Smith, B., Johnson, M., & Garcia, P. (2023). Semantic Technologies for Education and Training: State-of-the-Art and Challenges. *Computers & Education*, 180, 104584.

V. V. Prykhodniuk,
O. S. Kuzmenko,
V. V. Gorburokov

ONTOLOGY-DRIVEN PROCESSING OF STATISTICAL DATA FROM THE REPORTS OF REGIONAL OFFICES OF THE NC “JASU”

Abstract. The article presents the results of a study aimed at substantiating the conceptual and methodological foundations for the creation of an ontology-driven system for processing statistical data from the reports of regional branches of the National Center “Junior Academy of Sciences of Ukraine” (NC “JASU”). The relevance of the research is determined by the urgent need to automate the collection and processing of large volumes of educational data, which will contribute to increased transparency, accuracy, and analytical value of the reports for effective decision-making. Based on the analysis of current practices in collecting and structuring statistical reports within the NC “JASU”, key

problems were identified: lack of unified data structures, fragmentation of information, and difficulty in comparative analysis. The study proposes a concept for an ontology-driven system that ensures a formalized description of the domain, automated collection, semantic transformation, integration, and verification of statistical data. The methodological principles of the system's development are substantiated, including ontological modeling, interoperability, semantic consistency, and integration with digital educational services. Technological approaches to the implementation of the system are examined, as well as its effectiveness in enhancing the quality of educational analytics and strategic planning at the NC "JASU". The potential for scaling the proposed model to other educational and scientific institutions in Ukraine is outlined, with adaptation of ontologies to relevant domain-specific contexts. The results of the research are of practical value for the development of innovative educational analytics and digital governance in the scientific and educational environment.

Keywords: ontology-driven systems, educational statistics, Junior Academy of Sciences of Ukraine, digital transformation, semantic integration, educational analytics, strategic planning.

ІНФОРМАЦІЯ ПРО АВТОРІВ

Приходнюк Віталій Валерійович — канд. техн. наук, старший дослідник, завідувач відділу створення і використання інтелектуальних мережних інструментів, НЦ «Мала академія наук України», м. Київ, Україна, tangens91@gamdil.com; ORCID ID: <http://orcid.org/0000-0002-2108-7091>

Кузьменко Ольга Степанівна — д. пед. наук, професор, учений секретар секретаріату Вченої ради, Донецький державний університет внутрішніх справ, м. Кропивницький, Україна; провідний науковий співробітник відділу інформаційно-дидактичного моделювання, НЦ «Мала академія наук України», м. Київ, Україна, Kuzimenko12@gmail.com; ORCID ID: <http://orcid.org/0000-0003-4514-3032>

Горбурков В'ячеслав Вікторович — канд. техн. наук, старший дослідник, науковий співробітник відділу створення і використання інтелектуальних мережних інструментів, НЦ «Мала академія наук України», м. Київ, Україна, slavon07@gmail.com; ORCID ID: <http://orcid.org/0000-0002-2758-7724>

INFORMATION ABOUT THE AUTHORS

Prykhodniuk V. V. — PhD in Engineering, Senior Researcher, Head of the Department of Creation and Use of Intelligent Network Tools, NC "Junior Academy of Sciences of Ukraine", Kyiv, Ukraine, tangens91@gamdil.com; ORCID ID: <http://orcid.org/0000-0002-2108-7091>

Kuzmenko O. S. — D. Sc. in Pedagogy, Professor, Academic Secretary of the Secretariat of the Academic Council, Donetsk State University of Internal Affairs, Kropyvnytskyi, Ukraine; Leading researcher of the Department of Information and Didactic modelling, NC "Junior Academy of Sciences of Ukraine", Kyiv, Ukraine, Kuzimenko12@gmail.com; ORCID ID: <http://orcid.org/0000-0003-4514-3032>

Gorburukov V. V. — PhD in Engineering, Senior Researcher, Researcher at the Department of Creation and Use of Intelligent Network Tools, NC "Junior Academy of Sciences of Ukraine", Kyiv, Ukraine, slavon07@gmail.com; ORCID ID: <http://orcid.org/0000-0002-2758-7724>

Стаття надійшла до редакції / Received 30.05.2025