

УДК 81'322:004.8

DOI <https://doi.org/10.24919/2308-4863/83-3-35>**Володимир ПАСІЧНИК,***orcid.org/0000-0002-5231-6395**доктор технічних наук, професор,
професор кафедри інформаційних систем та мереж
Національного університету «Львівська політехніка»
(Львів, Україна) vrasichnyk@gmail.com***Максим ЯРОМИЧ,***orcid.org/0009-0005-3299-6695**аспірант кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»
(Львів, Україна) yamax0312@gmail.com*

ВЕЛИКІ МОВНІ МОДЕЛІ ТА ОНТОЛОГІЇ У ФІЛОЛОГІЧНИХ ДОСЛІДЖЕННЯХ: АНАЛІТИЧНИЙ ОГЛЯД ДЖЕРЕЛ

У статті досліджено можливість інтеграції великих мовних моделей (LLM) і онтологій для вирішення завдань у галузі філології. Підкреслено актуальність цієї теми у зв'язку з необхідністю автоматизації традиційних філологічних досліджень, які, попри зростання доступу до цифрових джерел, все ще виконуються здебільшого вручну. Впровадження новітніх технологій дозволяє значно підвищити ефективність аналізу текстів, таких як семантичний, стилістичний аналіз, класифікація за жанрами, генерація текстів, пошук міжтекстових зв'язків і створення підсумків. Розглянуто роль великих мовних моделей, що базуються на трансформерних архітектурах, таких як GPT, BERT та їх модифікації. Показано, як механізм самоуваги дозволяє моделям аналізувати текстові одиниці у широкому контексті, виявляючи складні семантичні зв'язки, стилістичні особливості та міжтекстові кореляції. Описано переваги LLM, зокрема автоматизацію складних аналітичних процесів, зниження витрат часу та ресурсів, а також високу точність опрацювання великих текстових баз. Особливий акцент зроблено на використанні онтологій для підвищення точності і структурованості досліджень. Вони забезпечують формалізоване представлення знань, що дозволяє знизити рівень неоднозначності при аналізі текстів. Крім того, онтології створюють основу для вирішення таких проблем LLM, як «галюцинації» – випадки генерації тексту, який виглядає правдоподібно, але є фактично неточним. У статті розглядаються конкретні приклади застосування LLM та онтологій у таких напрямках, як семантичний, стилістичний та лексикографічний аналіз, класифікація текстів, пошук міжтекстових зв'язків та генерація тексту. Підкреслено важливість розробки спеціалізованих онтологій, які враховують особливості мови, культури та літератури. Такі онтології можуть застосовуватися для міждисциплінарних досліджень, автоматизації створення словників і навчальних матеріалів, а також для поглибленого вивчення мовної та культурної еволюції. Сформовано висновки про значний потенціал LLM та онтологій для масштабування філологічних досліджень, їх автоматизації та підвищення точності. Обґрунтовано перспективи подальших досліджень у напрямку покращення інтеграції цих технологій, зокрема шляхом створення адаптивних онтологій для специфічних завдань і розроблення інструментів для зменшення помилок моделей. Такий підхід дозволяє відкрити нові горизонти для філології, забезпечуючи якісно новий рівень досліджень у цій галузі.

Ключові слова: великі мовні моделі, онтології, семантичний аналіз, стилістичний аналіз, генерація тексту.

Volodymyr PASICHNYK,*orcid.org/0000-0002-5231-6395**Doctor of Technical Sciences, Professor,
Professor at the Department of Information Systems and Networks
Lviv Polytechnic National University
(Lviv, Ukraine) vrasichnyk@gmail.com***Maksym YAROMYCH,***orcid.org/0009-0005-3299-6695**PhD student at the Department of Applied Linguistics
Lviv Polytechnic National University
(Lviv, Ukraine) yamax0312@gmail.com*

LARGE LANGUAGE MODELS AND ONTOLOGIES IN PHILOLOGICAL RESEARCH: ANALYTICAL REVIEW OF SOURCES

The article explores the possibility of integrating large language models (LLMs) and ontologies to address challenges in the field of philology. The relevance of this topic is emphasized due to the need to automate traditional philological research, which, despite increasing access to digital resources, is still largely conducted manually. The implementation of advanced technologies significantly enhances the efficiency of text analysis, such as semantic and stylistic analysis, genre classification, text generation, intertextual link discovery, and summarization. The study examines the role of large language models based on transformer architectures like GPT, BERT, and their modifications. It demonstrates how the self-attention mechanism enables these models to analyze textual units in a broader context, revealing complex semantic relationships, stylistic features, and intertextual correlations. The advantages of LLMs are highlighted, including the automation of complex analytical processes, reduction of time and resource costs, and high accuracy in processing large text corpora. Special attention is given to the use of ontologies to improve the precision and structure of research. Ontologies provide a formalized representation of knowledge, which reduces ambiguity in text analysis. Furthermore, they form the basis for addressing challenges like «hallucinations» in LLMs – instances where generated text appears plausible but is factually inaccurate.

The article discusses specific applications of LLMs and ontologies in areas such as semantic, stylistic, and lexicographic analysis, text classification, intertextual link discovery, and text generation. The importance of developing specialized ontologies that consider linguistic, cultural, and literary features is underscored. These ontologies can be used for interdisciplinary research, automation of dictionary creation, educational material development, and in-depth studies of linguistic and cultural evolution. Conclusions are drawn about the significant potential of LLMs and ontologies to scale philological research, automate processes, and improve accuracy. The prospects for further research are substantiated, particularly in improving the integration of these technologies through the creation of adaptive ontologies for specific tasks and developing tools to reduce model errors. This approach paves the way for new horizons in philology, ensuring a qualitatively new level of research in this field.

Key words: large language models, ontologies, semantic analysis, stylistic analysis, text generation.

Постановка проблеми. Філологічні дослідження, на відміну від багатьох галузей, станом на сьогодні все ще проводяться традиційними способами – читанням книг тих чи інших авторів, виділенням основних ідей та проведенням порівнянь у ручному режимі. І якщо в аспекті доступу до джерел уже використовуються електронні ресурси та Інтернет, то в плані проведення аналізу досі вдаються до підходів, характерних для минулого століття.

Зрозуміло, що такий аналіз текстів є дуже часозатратним, вимагає великих ресурсів та призводить до того, що філологія розвивається не так швидко, як могла б із використанням сучасних інформаційних технологій. Одним із найбільш перспективних способів подолання цієї проблеми є використання сучасних великих мовних моделей (LLM), зокрема на основі онтологічного підходу.

До ключових переваг використання LLM у філологічних дослідженнях належать автоматизація семантичного та стилістичного аналізу тексту, автоматичне виділення концептів та інтертекстуальних зв'язків, жанрова класифікація та розподіл текстів за тематикою. Також великі мовні моделі показали свою ефективність у ідентифікації емоційного забарвлення текстів. Онтологічний підхід при цьому є корисним при побудові обширних баз даних із великої кількості творів та сприяє оптимізації автоматизованого аналізу. Ще одним важливим прикладом застосування LLM у філологічних

дослідженнях є генерація текстів, таких як літературні огляди або матеріали для представлення граматичних конструкцій.

Варто зазначити, що при використанні великих мовних моделей постають і певні виклики, зокрема проблема “галюцинацій”, коли модель помилково утворює хибні зв'язки та може видавати неправильні результати, а також нестача структурованої інформації, у тому числі спеціалізованих онтологій у галузі філології. Також, недостатньо досліджень проведено на тему інтеграції онтологій із великими мовними моделями.

Таким чином, необхідно досліджувати специфіку та ефективність використання великих мовних моделей у філології, адже це відкриває нові, раніше недоступні, можливості для автоматизації традиційних завдань філології, зокрема глибшого аналізу мовних явищ, підвищення точності текстової класифікації та генерування змістовних текстів у різних контекстах. Це має величезний потенціал для підвищення ефективності та масштабування досліджень.

Аналіз досліджень. Тематика використання великих мовних моделей та онтологій у філологічних дослідженнях є порівняно новою, але стрімко розвивається завдяки прогресу в галузі опрацювання природної мови (NLP) та штучного інтелекту. На сьогодні у науковій літературі накопичено значну кількість даних щодо потенціалу та викликів, пов'язаних із впровадженням цих

технологій, але ряд питань залишається відкритими.

Одним із ключових напрямків досліджень є семантичний аналіз текстів. Великі мовні моделі, такі як GPT-4 та BERT, продемонстрували високу ефективність у виявленні значень слів і фраз у контексті (Mukanova et al., 2024). Водночас важливу роль відіграють онтології, які додають структурованість у семантичному аналізі (Zulkipli et al., 2022; Giglou et al., 2023). Проте залишається відкритим питання інтеграції онтологій із LLM для роботи з багатозначними словами, культурно специфічними поняттями та складними семантичними структурами.

Стилістичний аналіз також привертає значну увагу дослідників. LLM виявляють стилістичні маркери текстів із точністю до 98%, як продемонстровано у дослідженні (Oraa, 2024). Їх застосування включає визначення авторства, аналіз стилістичних фігур, таких як метафори та іронія, а також класифікацію текстів за жанрами та стилями. Однак проблема виявлення контекстуальних стилістичних фігур, таких як алюзії, залишається актуальною, адже навіть сучасні моделі стикаються з труднощами через складність культурних і літературних контекстів.

У сфері класифікації текстів великі мовні моделі демонструють суттєву перевагу над традиційними методами (Giglou et al., 2023). Онтології при цьому сприяють формалізації зв'язків між текстовими категоріями, що дозволяє зменшити рівень семантичної неоднозначності. Водночас викликом є недостатня деталізація онтологій для гібридних жанрів або складних текстових категорій.

Генерація тексту є одним із найважливіших напрямків застосувань LLM у філології. Моделі, такі як GPT-4, можуть створювати текст із високою граматичною, стилістичною та семантичною точністю. Це відкриває можливості для автоматизації створення навчальних матеріалів, наукових текстів, літературних оглядів і навіть художніх творів (To et al., 2024). У поєднанні з онтологіями LLM здатні враховувати специфіку історичного чи культурного контексту, створюючи адаптовані матеріали. Водночас проблема забезпечення фактичної точності та узгодженості стилю у генерованих текстах залишається актуальною, особливо для спеціалізованих текстів.

Ще одним перспективним напрямком є пошук міжтекстових зв'язків, де LLM у поєднанні з онтологіями можуть аналізувати інтертекстуальні посилання, алюзії та ремінісценції (Umphrey et al., 2024). Проте викликом залишається виявлення

глибинних алюзій і прихованих текстових зв'язків у багатокультурному контексті.

Проблема «галюцинацій» у LLM, тобто генерації текстів, які є фактично неточними, є ще одним важливим аспектом досліджень. Як зазначено у роботах (Nananukula et al., 2024; Agrawal et al., 2023), використання онтологій допомагає зменшити ризик таких помилок, забезпечуючи доступ до структурованої та перевіреної інформації. Графи знань, інтегровані з LLM, можуть суттєво покращити якість текстів, але проблема повної усунення «галюцинацій» залишається відкритою.

Таким чином, основні досягнення в галузі досліджень LLM та онтологій включають автоматизацію семантичного та стилістичного аналізу, точну класифікацію текстів, виявлення міжтекстових зв'язків та генерацію контекстуально адаптованих текстів. Водночас актуальними залишаються виклики, пов'язані з адаптацією моделей до специфічних культурних і лінгвістичних контекстів, зменшенням «галюцинацій» та розвитком спеціалізованих онтологій для більш точного аналізу текстів. Подальші дослідження можуть бути спрямовані на інтеграцію цих технологій для вирішення складних філологічних задач, що відкриває нові горизонти для гуманітарних наук.

Мета статті – дослідити можливості використання великих мовних моделей та онтологій для проведення досліджень у галузі філології, а також показати їх ефективність для автоматизації дослідницьких процесів.

Виклад основного матеріалу. Великі мовні моделі стали важливим етапом у розвитку лпрацювання природної мови (Natural Language Processing), починаючи з появи нейронних мереж, і з поступовим ускладненням архітектур їх роль лише зростала. Перші моделі, такі як Word2Vec, створювали векторні представлення слів, що дозволяло розуміти їхнє значення через контекст. Однак революція почалася з появи трансформерів (Vaswani et al., 2017), архітектури, яка замінила традиційні рекурентні та згорткові нейронні мережі, що мали обмеження у швидкості та здатності працювати з довгими текстами. Ключовою інновацією трансформерів став механізм *self-attention* (самоуваги), який дозволяє моделі визначати, які частини тексту є найбільш важливими для аналізу кожного слова чи фрази. Цей підхід дозволив ефективно працювати з послідовностями тексту будь-якої довжини та заклав основу для моделей, орієнтованих на розуміння та генерацію текстів, таких як, зокрема, BERT, RoBERTa та GPT.

У цій статті розглядаються можливості використання великих мовних моделей та онтологій

в аспекті таких філологічних задач, як: семантичний аналіз, стилістичний аналіз, класифікація текстів, лексикографічний аналіз, пошук інформації у текстових корпусах, аналіз міжтекстових зв'язків, генерація тексту та створення підсумків, створення освітніх матеріалів, а також обґрунтовується необхідність побудови спеціалізованих філологічних онтологій. Крім цього, описано проблему «галюцинацій» та способи боротьби з нею за допомогою онтологічного підходу.

Важко переоцінити також роль великих мовних моделей у збагаченні онтологій, що сприяє підвищенню рівня структуризації знань (Giglou et al., 2023). Однією із основних переваг великих мовних моделей є здатність виявляти глибинні закономірності у великих обсягах текстів. Наприклад, моделі GPT-3 і GPT-4 тренуються на корпусах даних, які охоплюють мільярди текстових документів, завдяки чому вони можуть не лише відтворювати текст у різних стилях, а й відповідати на складні запити, враховуючи при цьому контекст. Завдяки цьому такі моделі зручно використовувати для розв'язання широкого спектру філологічних задач. Зокрема, показано ефективність того, як LLM можуть бути інтегровані в роботу з онтологіями для семантичного аналізу великих текстових баз, забезпечуючи при цьому гнучкість і адаптивність (Mukanova et al., 2024).

Поширення LLM у наукових і прикладних дослідженнях значною мірою сприяло їхній здатності до генералізації, внаслідок чого вони здатні виявляти не лише приховані зв'язки між концептами, а також допомагати у створенні нових знань (Neuhaus, 2024). Це робить їх активним учасником у розвитку гуманітарних наук, зокрема філології.

Семантичний аналіз. Розглянемо основні класи філологічних задач, які доцільно вирішувати за допомогою онтологій та великих мовних моделей. Однією із таких задач є семантичний аналіз, спрямований на дослідження значень слів, фраз та їхніх зв'язків у контексті. Ця задача вимагає не лише врахування лексичних значень, а й розуміння багаторівневих семантичних структур, що виникають у текстах. Інтеграція великих мовних моделей та онтологій надає інструменти для більш глибокого аналізу та точніших результатів. Оскільки великі мовні моделі, такі як GPT чи BERT, навчаються на багатомовних текстових корпусах, то вони можуть виявляти значення слова, враховуючи його контекст, навіть якщо це слово має декілька значень. Наприклад, у дослідженні (Mukanova et al., 2024) показано, як LLM використовуються

для семантичного узагальнення, що дозволяє ідентифікувати загальні та унікальні властивості текстів. У результаті було встановлено, що модель успішно виділяє ключові поняття, знаходить їхні семантичні зв'язки та автоматично класифікує концепти, у таких тестових завданнях досягнуто точності понад 85%.

Також LLM ефективно працюють із фразеологізмами, ідіомами та іншими складними мовними конструкціями, які традиційно створювали труднощі для автоматичного аналізу текстів. Крім цього, вони можуть також використовуватися для семантичного пошуку – замість пошуку точного збігу слів, вони враховують семантичну схожість текстових одиниць. Це дозволяє знайти тексти, які мають схожі значення, навіть якщо в них використовуються інші слова. Для цього, наприклад, було розроблено систему OntoChatGPT, яка поєднує LLM із семантичними запитами, забезпечуючи точний пошук у великих текстових базах. Запити, згенеровані на основі онтологій, дозволяли отримувати відповіді, які були на 30% точнішими порівняно зі звичайними запитами без онтологічної підтримки. Онтології, у свою чергу, забезпечують структурованість і чіткість у семантичному аналізі шляхом створення формальних моделей зв'язків між концептами, що дозволяє уникати двозначності при інтерпретації значень. У роботі (Zulkipli et al., 2022) показано, як онтології використовуються для семантичного узагальнення текстів, забезпечуючи систематичний підхід до класифікації концептів. У дослідженні (Palagin et al., 2023) наведено приклад використання онтологій для семантичного аналізу великих текстів та продемонстровано, як онтологічні моделі можуть бути використані для виявлення семантичних та лексичних закономірностей. Автори запропонували інформаційну систему OntoChatGPT, яка поєднує онтологічний підхід та застосування найпопулярнішої на сьогодні великої мовної моделі ChatGPT.

Особливо корисні онтології у задачах виявлення семантичних зв'язків між текстовими одиницями, так, у (Yorsh et al., 2022) описано метод використання онтологій для побудови зв'язків між термінами в текстах, що є основою для створення семантичних графів. Такі графи допомагають у визначенні зв'язків між словами та фразами, а також побудові моделей їхньої взаємодії в межах складних текстів. Експерименти показали, що точність автоматичного збагачення онтологій досягала 90%, а інтеграція онтологій із LLM дозволяла майже вдвічі зменшити кількість помилок у семантичному аналізі. Онтології об'єднують кон-

цепти та зв'язки між ними в ієрархічну структуру, тому вони є зручними для опису складних мовних явищ, зокрема семантичних зв'язків, синтаксичних конструкцій та стилістичних особливостей. Це значно спрощує аналіз літературних творів, а також пошук міжтекстових зв'язків (Giglou et al., 2023; Neuhaus, 2024).

Наразі актуальними залишаються проблеми адаптації моделей до багатозначних слів і культурно специфічних понять, а також опрацювання глобальних семантичних зв'язків у довгих текстах. Крім того, автоматизована побудова онтологій для рідковживаних мов потребує удосконалення через обмеженість навчальних даних.

Стилістичний аналіз. Використання LLM та онтологій відкриває також багато можливостей для автоматизації стилістичного аналізу. Великі мовні моделі здатні розпізнавати стилістичні особливості текстів завдяки тренуванню на великих корпусах даних. У дослідженні (Орага, 2024) описано, як LLM використовуються для стилістичного аналізу. Автори виявили, що такі моделі значно перевершують традиційні методи у точності визначення стилістичних маркерів і класифікації текстів за стилем. Так, запропонована модель StyloAI змогла ідентифікувати авторство освітніх текстів з точністю до 98%, що значно перевищує показники попередніх моделей. У статті (Калініченко, 2023) розглянуто різні типи семантичних моделей, такі як Latent Semantic Analysis (LSA), тематичне моделювання, розпізнавання іменованих сутностей (NER), аналіз настроїв та синтаксичний розбір, які використовуються для аналізу текстів. LLM також можуть визначати тональність тексту, рівень формальності та навіть жанрові особливості. Завдяки механізму самоуваги ці моделі здатні аналізувати стилістичні фігури, такі як метафори, алегорії чи ідіоми, враховуючи широкий контекст. При цьому, використання онтологій дозволяє формалізувати знання про стилістичні категорії, їх функції та взаємозв'язки, створюючи основу для систематичного аналізу текстів. При такому підході LLM надають гнучкість і здатність до розуміння складних контекстів, а онтології забезпечують структурованість даних і перевіреність результатів.

Перспективи подальших досліджень у цьому напрямку вбачаємо у виявленні складних стилістичних фігур, таких як іронія чи алюзії, через їхню контекстуальну природу. Також онтології часто не охоплюють повного спектра стилістичних категорій, що обмежує їхню інтеграцію з мовними моделями.

Класифікація текстів. Ще одним класом філологічних задач, де великі мовні моделі та онтології продемонстрували свою ефективність, є класифікація текстів, що має на меті автоматичне визначення жанру, стилю, тематики або інших характеристик тексту. LLM забезпечують високу ефективність у завданнях класифікації завдяки їх здатності аналізувати складні патерни в текстах, враховуючи як локальні, так і глобальні контекстуальні зв'язки в тексті. Це робить їх особливо корисними для аналізу багатозначних або контекстно залежних термінів. Інтеграція великих мовних моделей та онтологій забезпечує високий рівень точності та гнучкості, що робить їх незамінними інструментами для цього завдання. У дослідженні (Giglou et al., 2023) було показано, що LLM досягають точності у 94% при класифікації текстів, що суттєво перевищує результати традиційних засобів, причому модель GPT продемонструвала здатність правильно виділяти жанри та стилі навіть за умови неоднозначності або неповноти даних. Онтології, у свою чергу, формалізують зв'язки між концептами, створюючи ієрархічні моделі, які визначають взаємозв'язки між різними категоріями текстів. У роботі (Dessi et al., 2020) зазначено, що використання онтологій у побудові текстових категорій дозволяє знизити рівень семантичних неоднозначностей і підвищити точність класифікації на 15%. Дослідження (Kowsari et al., 2019) пропонує огляд сучасних алгоритмів класифікації текстів, включаючи класичні методи машинного навчання та нейронні мережі.

Крім цього, онтології сприяють підвищенню інтерпретованості результатів, що важливо для гуманітарних наук, де точність класифікації не менш важлива, ніж зрозумілість критеріїв, на основі яких вона здійснюється. Такий підхід є, безумовно, перспективним для завдань, пов'язаних із класифікацією літературних творів, організацією текстових баз даних і створенням рекомендаційних систем. Наразі класифікація текстів за допомогою LLM і онтологій стикається з проблемами узгодження семантики між різними жанрами та стилями, що призводить до неоднозначностей. Іншою проблемою є недостатня деталізація онтологій для складних текстових категорій, таких як гібридні жанри. Подальші дослідження можуть бути зосереджені на створенні спеціалізованих онтологій і адаптації LLM для розпізнавання багаторівневих жанрових структур.

Лексикографічний аналіз. Одним із найважливіших напрямків у сучасній лінгвістиці є лексикографічний аналіз. Він забезпечує фіксацію

лексичних одиниць, а також систематизацію знань про їхні значення, синтаксичні функції, стилістичні особливості та міжмовні відповідності. Підходи, що інтегрують великі мовні моделі та онтології, дозволяють суттєво автоматизувати цей процес, підвищуючи його ефективність, точність і глибину аналізу. Моделі, побудовані на трансформерній архітектурі, такі як GPT або BERT, демонструють високу ефективність у опрацюванні та аналізі текстів. Завдяки великим обсягам текстів, на яких вони навчаються, вони можуть визначати значення слів у контексті, враховуючи їхню полісемію, синонімію та конотативні відтінки. Це особливо важливо для лексикографічного аналізу, де потрібно не лише виявити слово, але й виокремити його різноманітні аспекти. У роботі (Mukanova et al., 2024) продемонстровано, як LLM можуть автоматично витягувати лексичні одиниці, їхні визначення та приклади використання з великих текстових баз. Автори зазначають, що використання LLM у таких завданнях дозволило значно скоротити час, витрачений на створення словникових статей, і забезпечило їхню відповідність сучасним мовним тенденціям. Моделі також ефективно виявляють нові слова або фрази, які щойно увійшли в активний вжиток, але ще не зафіксовані у традиційних словниках. При цьому, великі мовні моделі здатні проводити частотний аналіз лексем і визначати їхній стильовий або жанровий контекст. Це дозволяє створювати словники, що враховують не лише базові значення слів, але й їхні функції у специфічних стилях, таких як науковий чи художній текст. За допомогою онтологій зручно представляти складні семантичні зв'язки між словами, такі як синонімія, антонімія, гіперонімія та гіпонімія, що дає змогу формалізувати знання та створювати структуровані моделі мови.

Автори роботи (Михайлюк, 2012) пропонують методологію інтеграції онтологій у лексикографію та демонструють, що використання онтологій дозволяє не лише автоматизувати процес створення словників, але й забезпечити більшу точність у визначенні семантичних відношень між словами. Наприклад, у випадках полісемії онтології допомагають чітко розмежувати різні значення слова залежно від контексту. Корисними є онтології також для аналізу метамовних структур, що особливо важливо у випадках міжмовної лексикографії. Вони дозволяють встановлювати відповідності між лексемами різних мов, зважаючи на культурні й семантичні відмінності. Використання LLM та онтологій дає змогу також створювати динамічні словники, які оновлюються в реальному часі залежно від змін у мовному

середовищі, що є важливим для спеціалізованих сфер, де лексика швидко змінюється, наприклад, у технічній чи науковій комунікації. У дослідженні (Oruganti et al, 2023) описано систему, яка використовує LLM для автоматичного поповнення семантичного лексикону інтелектуального агента новими термінами. Цей підхід дозволяє агенту адаптуватися до нових мовних явищ без втручання людини, що є ключовим для підтримки актуальності словника в динамічних галузях. Прикладом може слугувати також платформа АВВУУ Lingvo Live, яка пропонує «живий» словник, що постійно оновлюється користувачами та експертами, забезпечуючи актуальність лексичних даних. Цей сервіс поєднує авторитетні словники з народним контентом, що дозволяє швидко реагувати на появу нових термінів та значень.

Викликами залишається адаптація до швидких змін у сучасній лексиці, зокрема науковій та сленговій, а також недостатня автоматизація процесу виділення семантичних зв'язків між словами в рідковживаних мовах. Вирішити ці проблеми може допомогти вдосконалення методів збагачення онтологій та інтеграції LLM для більш точного і швидкого аналізу.

Пошук у текстових корпусах. Високу ефективність демонструють LLM та онтології у пошуку інформації у великих корпусах, що є невід'ємною частиною аналізу текстових даних. Завдяки інтеграції великих мовних моделей та онтологій цей процес став значно швидшим та точнішим. Останні версії LLM, зокрема GPT-4, здатні розуміти складні запити природною мовою, що дозволяє використовувати їх як інструменти для семантично збагаченого пошуку. Вони не лише аналізують синтаксис запиту, а й виявляють контекстуальні взаємозв'язки між словами. У роботі (Mukanova et al., 2024) продемонстровано, як GPT може бути налаштовано для пошуку релевантної інформації у великих корпусах текстів; при цьому точність результатів значно перевищувала отримані традиційними засобами пошуку. Також LLM успішно застосовуються для пошуку у спеціалізованих текстових базах, таких як наукові статті чи літературні твори. Наприклад, у базах літературних текстів вони можуть ідентифікувати повторювані теми, ключові концепти або стилістичні прийоми, що робить їх незамінними в літературознавчих дослідженнях. Використання онтологічного підходу при цьому дозволяє забезпечити багатомовність у пошукових запитах, що особливо важливо для аналізу текстів у глобальному контексті. Через онтологічне відображення понять одного й того ж значення в

різних мовах можна ефективно виконувати пошук у мультигенераційних текстових базах.

Пошук у текстових корпусах із використанням LLM і онтологій наразі стикається з такими проблемами, як врахування контекстуальних значень слів і фраз, а також недостатньою деталізацією онтологій для специфічних предметних галузей, що може призводити до неповноти або хибності результатів. Перспективними є дослідження, спрямовані на вдосконалення моделей для обробки контексту та розширення онтологій для забезпечення більш структурованого пошуку.

Пошук міжтекстових зв'язків. В аспекті текстового аналізу важливим завданням є пошук міжтекстових зв'язків, із чим великі мовні моделі та онтології також мають великий потенціал допомогти. LLM можуть ефективно аналізувати запозичення ідей, сюжетів, образів або стилістичних прийомів, відслідковуючи застосування цитувань, ремінісценцій, пародій, алюзій та інтертекстуальних посилань. При опрацюванні великих обсягів тексту вони виявляють неочевидні зв'язки між творами, які раніше могли залишатися непоміченими, а також можуть автоматично визначати теми, мотиви чи стилістичні особливості, що об'єднують різні твори, що робить їх корисними для порівняльного аналізу. Використання онтологій при підході до цієї задачі підвищує рівень чіткості та систематизації завдяки структурованому представленню зв'язків між текстовими елементами, а також дозволяє зменшити кількість хибних результатів. У роботі (Umphrey et al., 2024) досліджується здатність LLM виявляти інтертекстуальні зв'язки в давньогрецьких текстах. Результати показують, що моделі можуть ідентифікувати прямі цитати, алюзії та відлуння між текстами, відкриваючи нові перспективи для вивчення літературних взаємодій. Застосування LLM та онтологій також надає можливість такого потужного інструментом для візуалізації, як інтерактивні карти інтертекстуальності, що відкриває нові можливості для досліджень у даному напрямку.

У пошуку міжтекстових зв'язків можна виділити низку актуальних проблем, зокрема виявлення глибинних алюзій, прихованих цитувань та інтертекстуальних посилань, особливо у великих корпусах текстів. Також досить складно інтегрувати онтології, які б адекватно відображали складні зв'язки між текстами різних культур і епох. Перспективу подальших досліджень вбачаємо у розробці більш адаптивних онтологій та вдосконаленні моделей для виявлення контекстуальних і семантичних взаємозв'язків.

Генерація тексту. Основною функцією великих мовних моделей є генерація тексту, що, поза сумнівом, має величезне значення та широке застосування у філологічних дослідженнях. Завдяки своїй здатності створювати зв'язний, контекстуально відповідний текст, LLM стають потужними інструментами для створення художніх та наукових текстів, реконструкції втрачених текстів, генерації описів, анотацій, перекладів та ще цілого ряду завдань, які потребують написання тексту.

Трансформерна архітектура, що лежить в основі більшості сучасних великих мовних моделей, таких як GPT-4, дозволяє використовувати механізм самоуваги для аналізу контексту слів у тексті, що дозволяє передбачати наступне слово або фразу з високою точністю. Цей підхід забезпечує створення текстів, які не лише мають граматичну коректність, але й враховують стилістичні, семантичні та контекстуальні аспекти. У дослідженні (To et al., 2024) продемонстровано, що LLM здатні адаптуватися до різних стилів написання, включаючи науковий, художній або публіцистичний, що робить їх надзвичайно корисними для вивчення стилістичних характеристик текстів. Генерація текстів у стилі певного автора дозволяє дослідникам ефективніше вивчати стилістичні особливості його творів. Для прикладу, можна використати імітації стилю Шекспіра для дослідження того, як стилістичні елементи впливають на сприйняття тексту. У поєднанні з онтологіями LLM здатні генерувати тексти, що імітують історичний контекст, створюючи історично точні матеріали, що є корисним, наприклад, при визначенні автентичності тих чи інших історичних текстів. Так, автори статті (Schimmenti et al., 2024) розробили методологію, яка використовує LLM для вибору, вилучення та класифікації релевантних тверджень щодо автентичності документів без потреби в додатковому навчанні. Отримані дані структуруються за допомогою технологій семантичної павутини, що дозволяє створювати каталоги документів із зазначенням дискусій про їхню автентичність. Великі мовні моделі також є ефективними для моделювання мовних змін, тобто дозволяють досліджувати, як сучасні тексти могли б виглядати в стилі минулих епох, імітуючи мовні структури та вирази, властиві тим часам, або ж навпаки – історичні тексти у сучасному контексті.

У дослідженні (Dessi et al., 2020) автори описали, як LLM у поєднанні з онтологіями здатні генерувати тексти, які точно відображають зміст знань у певній галузі. Наприклад, система змогла створити детальні описи наукових концептів, які

використовуються в освітніх матеріалах. Розглянуто використання методів обробки природної мови (NLP) та машинного навчання для створення графів знань у наукових текстах. Це дозволяє структурувати знання, витягнуте з текстових джерел, і створювати інтерактивні бази знань, які можуть бути корисними для філологічних завдань, наприклад, аналізу літературного контексту. У роботі (Giglou et al., 2024b) продемонстровано, що LLM можуть генерувати текстові описи для онтологічних концептів, допомагаючи автоматизувати процес документування онтологій із використанням запропонованого фреймворку LLMs4OM. Представлені у (Goel et al. 2021) сучасні методи керованої генерації текстів із акцентом на наукові тексти дозволяють модифікувати стиль, тон і структуру генерованого тексту залежно від завдань, що є корисним для створення навчальних матеріалів у філології. Зручні онтології і для отримання знань із наукових робіт. Зокрема, робота (Hu et al., 2024) пропонує інноваційний підхід до автоматизації витягування знань зі спеціалізованої наукової літератури, який базується на інтеграції великих мовних моделей та онтологій. Розроблена система з прогресивним онтологічним налаштуванням дозволяє підвищити точність результатів, що може бути адаптовано до гуманітарних наук для аналізу текстів.

Генерація тексту за допомогою LLM і онтологій стикається з проблемами забезпечення фактичної точності, узгодженості стилю та врахування складних контекстуальних особливостей. Важливою залишається також задача адаптації моделей до специфічних жанрових і тематичних вимог, особливо для спеціалізованих текстів. Для вирішення цих викликів необхідно вдосконалювати взаємодію між онтологіями та LLM, що дозволить створювати текстів із вищим рівнем релевантності та семантичної глибини.

Автоматичне створення підсумків. В аспекті генерації текстів доцільно згадати і про задачу автоматичного створення підсумків, у якій великі мовні моделі та онтології теж дуже потужно себе зарекомендували. Цей процес спрямований на зменшення обсягу тексту при збереженні ключової інформації та семантичного значення. Використання LLM у поєднанні з онтологіями відкриває нові можливості для створення якісних, точних і релевантних підсумків. Завдяки згаданій вище трансформерній архітектурі, яка забезпечує глибоке розуміння контексту, великі мовні моделі, такі як GPT-4, демонструють високу ефективність у генерації підсумків. Це дозволяє моделі аналізувати текст із врахуванням семантичних

зв'язків між словами, реченнями та абзацами. У роботі (To et al., 2024) зазначено, що LLM здатні адаптувати підсумки під різні стилі написання та цільові аудиторії, наприклад, у наукових текстах вони можуть виділяти найважливіші концепти, зберігаючи при цьому складну термінологію та лексичну точність. Відчутною перевагою LLM є також здатність працювати з великими обсягами текстів, швидко витягуючи ключову інформацію. Завдяки механізму самоуваги, такі моделі здатні розпізнавати релевантні елементи навіть у складних документах. Це робить їх незамінними для обробки спеціалізованих текстів, таких як, зокрема, медичні звіти або юридичні документи. Так, у юридичних додатках LLM вже використовуються для автоматичного створення резюме судових справ, допомагаючи юристам ефективніше аналізувати інформацію, тоді як у медичних дослідженнях вони створюють резюме наукових статей, що прискорює доступ лікарів до критично важливої інформації.

Онтології, зі свого боку, забезпечують формалізоване представлення знань, що дозволяє структурувати інформацію та виділяти ключові поняття, тому вони є особливо важливими для семантичної інтеграції даних. У роботі (Mukanova et al., 2024) описано, як онтології додають контекст і точність до текстів, дозволяючи системам уникати втрати важливих деталей; при цьому онтологічні моделі допомагають формалізувати відношення між поняттями, що є дуже важливим для створення змістовних підсумків.

Поєднання LLM та онтологій створює синергію, яка забезпечує важливі переваги для підсумовування текстів. Завдяки використанню онтологій для визначення ключових концептів, мовні моделі можуть зосередитися на релевантній інформації, ігноруючи несуттєві деталі, що значно підвищує точність отриманих результатів. Крім того, онтології забезпечують узгодженість термінів і структур, сприяючи збереженню семантичної цілісності тексту, особливо в складних міждомених аналізах. Інтеграція онтологій дозволяє моделям адаптувати підсумки до специфіки контексту, враховуючи особливості вихідного матеріалу, що забезпечує високу релевантність та відповідність кінцевого результату потребам користувача.

У дослідженні (Toro et al., 2024) підкреслюється, що використання онтологій разом із LLM дозволяє зберігати семантичну цілісність текстів навіть у випадках великого обсягу даних. Автори демонструють, що така інтеграція дозволяє знизити частоту помилкових або нерелевантних узагаль-

нень. У філології, де дослідження часто зосереджені на міжтекстових зв'язках, стилістичних особливостях і семантичних залежностях, LLM та онтології можуть автоматизувати рутинні процеси та поглиблювати розуміння тексту. Вони можуть бути застосовані для створення узагальнень, що враховують специфіку літературного аналізу, наприклад, для виявлення основних тем у творах певного автора або жанру, а також здатні підсумовувати велику кількість текстів, визначаючи спільні стилістичні елементи чи лексичні особливості. Завдяки онтологіям ці процеси стають більш структурованими: наприклад, стає можливим визначити зв'язки між лексичними групами, жанровими характеристиками або контекстуальними аспектами тексту.

Особливу цінність такі інструменти мають у вивченні еволюції літературних стилів. Підсумовування текстів, що створюються LLM з урахуванням онтологічних моделей, дозволяє філологам визначати зміну мовних тенденцій у творах різних епох, аналізувати вплив культурного контексту на формування літературних жанрів і стилів. Наприклад, онтології можуть допомогти систематизувати концепти, що стосуються романтизму чи модернізму, а LLM – згенерувати стислий, але змістовний опис ключових характеристик цих літературних напрямів. Крім того, LLM здатні автоматично створювати підсумки літературних творів різними мовами, виявляючи спільні та відмінні риси у стилістиці чи змісті, що робить їх корисними для проведення порівняльних літературних досліджень. При цьому онтологічний підхід забезпечує міжмовну семантичну узгодженість та збереження точності при аналізі перекладів чи адаптацій.

Таким чином, використання великих мовних моделей у поєднанні з онтологіями стає важливим інструментом для автоматичного підсумовування текстів. Цей підхід не лише підвищує точність та якість згенерованих підсумків, але й розширює межі використання цих технологій у різних галузях знань. Інтеграція LLM та онтологій сприяє покращенню доступу до знань, автоматизації рутинних завдань і створенню нових інструментів для роботи з текстами.

До викликів, які постають перед використанням LLM та онтологій для автоматичної генерації підсумків, можна віднести збереження семантичної точності, уникнення втрати ключової інформації та уникнення дублювань у тексті. Крім цього, актуальною проблемою залишається адаптація підсумків до специфіки філологічних досліджень, зокрема для складних літературних

або лінгвістичних текстів. Потенційним шляхом їх вирішення є вдосконалення механізмів інтеграції онтологій для покращення узгодженості та значущості підсумків.

Створення освітніх матеріалів. Генерація текстів за допомогою LLM є потужним інструментом для створення освітніх матеріалів. Вони можуть генерувати навчальні тексти, тестові завдання або адаптації текстів для різних рівнів мовної компетенції. Наприклад, система може створювати короткі перекази творів для початкового рівня або аналіз складних текстів для поглибленого вивчення. У статті (Tzirides et al., 2023) досліджується застосування генеративного штучного інтелекту, зокрема LLM, для автоматизованого створення навчальних текстів, тестових завдань та адаптації матеріалів під різні рівні знань студентів. Автори також аналізують, як чатботи, що працюють на основі великих мовних моделей (С-LLM), можуть бути інтегровані в освітні практики. Зокрема, розглядається застосування С-LLM для автоматизованого огляду та оцінювання складних студентських робіт, що може значно полегшити роботу викладачів.

Також, моделі можуть підсумовувати ключові аспекти літературних творів для створення навчальних посібників або адаптацій текстів для різних рівнів читачів, тоді як онтології забезпечують точність і контекстуальність поданої інформації. Поєднання LLM та онтологій не лише прискорює роботу з текстами, але й забезпечує глибше розуміння семантичних і стилістичних взаємозв'язків, відкриваючи нові горизонти для філології. Робота (Tkachenko et al., 2021) розглядає розробку та використання лінгвістичних онтологій для інтелектуальних освітніх систем. Автори описують створення онтологій для моделювання граматичних структур і лексичних залежностей, які можуть сприяти автоматизації навчання мовам та покращувати взаємодію користувачів із навчальними платформами.

Створення освітніх матеріалів за допомогою LLM і онтологій стикається з проблемами адаптації матеріалів до різних рівнів мовної компетенції та забезпечення їхньої коректності й узгодженості з навчальними цілями, тому варто зосередитися на розвитку моделей, здатних створювати персоналізовані освітні матеріали з урахуванням індивідуальних потреб учнів. У перспективі ці технології можуть стати основою для створення інтерактивних платформ, де філологи отримають змогу автоматично аналізувати та підсумовувати тексти, збагачуючи свої дослідження новими інсайтами.

Проблема «галюцинацій». Незважаючи на численні переваги використання великих мовних моделей, у тому числі у філологічних дослідженнях, при роботі з ними можуть виникати певні обмеження. Зокрема, до них належать так звані «галюцинації». Галюцинації – це генерація тексту, який виглядає правдоподібно, але не відповідає дійсності. Вони виникають через кілька причин: обмеженість чи упередженість даних, на яких була навчена модель; недосконалість архітектури, яка працює з ймовірнісними прогнозами, а також відсутність механізмів для перевірки фактологічної достовірності створеного тексту. Ця проблема особливо актуальна у сферах, де точність і достовірність інформації є критично важливими, таких як, наприклад, наукові публікації, адже в них навіть невелика неточність здатна значно знизити достовірність отриманих результатів. Одним із найперспективніших підходів до боротьби з галюцинаціями є інтеграція онтологій, які забезпечують структуроване представлення концептів, їхніх властивостей та зв'язків між ними. Онтології не лише додають чіткість і точність, але й забезпечують LLM доступ до релевантної та перевіреної інформації. Наприклад, у роботі (Nanapukula et al., 2024) розроблено онтологію, яка допомагає аналізувати та класифікувати помилки, що виникають під час генерації текстів, надаючи детальний опис їхньої природи та причин. Окремим видом онтологій, що демонструє високий потенціал у зменшенні кількості галюцинацій, є графи знань (knowledge graphs). У дослідженні (Agrawal et al., 2023) продемонстровано, що інтеграція графів знань з LLM дозволяє підвищити якість генерації текстів завдяки семантичній узгодженості й доступу до верифікованих даних. Це особливо важливо для складних міждисциплінарних завдань, де модель працює з великими обсягами інформації. Також, у роботі (Sansford et al., 2024) представлено спеціалізовану методику оцінки галюцинацій у LLM на основі графів знань. Такий підхід дозволяє, окрім власне виправлення помилок, оптимізувати роботу моделей через інтеграцію структурованих даних. Використання графів знань забезпечує збереження логічної і семантичної цілісності текстів, зменшуючи ризик створення хибної інформації.

Онтології можуть відігравати подвійну роль у боротьбі з галюцинаціями. По-перше, вони допомагають фільтрувати некоректні інтерпретації текстів, які виникають через культурні або стилістичні відмінності. По-друге, онтології можуть бути використані для створення більш структурованих літературознавчих моделей,

які дозволяють глибше зрозуміти текстові взаємозв'язки та стилістичні особливості. Це дає можливість уникати помилок у складних аналітичних завданнях, таких як міжтекстовий аналіз чи аналіз авторських стилів. Завдяки цьому інтеграція онтологій у роботу LLM є ефективним засобом для зменшення галюцинацій, який не лише підвищує точність та достовірність створених текстів, але й відкриває нові можливості для розвитку інтелектуальних систем, що працюють з текстами. Онтології забезпечують моделі доступ до надійної бази знань, яка служить основою для перевірки результатів і вдосконалення процесу генерації, що робить їх важливим інструментом як для філологічних досліджень, так і для інших галузей.

Актуальність спеціалізованих онтологій. Надзвичайно важливими для філологічних досліджень є спеціалізовані онтології, оскільки вони забезпечують формалізований і структурований підхід до представлення знань про мову, літературу та текстові феномени. При дослідженні багатогранних аспектів мови й тексту, зокрема їхньої стилістики, семантики, синтаксису та культурного контексту, існує значна потреба в інструментах, здатних систематизувати ці складні дані. Спеціалізовані онтології надають таку можливість, забезпечуючи точне представлення ключових концептів і їхніх зв'язків, що значною мірою розширює можливості аналізу та автоматизації. Такі онтології широко використовуються і в інших галузях гуманітарних наук, так, дослідження (Nguyen et al., 2022) пропонує онтологічну модель для представлення знань, узятих із юридичних документів. Система була протестована на земельному праві В'єтнаму та показала ефективність у визначення термінології та описі адміністративних процедур, пов'язаних із земельним правом, таким чином значно спростила роботу користувачів.

Особливістю спеціалізованих онтологій є їхня здатність адаптуватися до конкретних потреб дослідницького середовища. У літературознавстві вони можуть бути використані, зокрема, для опису тем, мотивів, образів і стилістичних прийомів, які зустрічаються у творах певного автора або літературного жанру. Це дозволяє не лише автоматизувати процеси класифікації текстів, але й проводити глибокий аналіз міжтекстових зв'язків і впливів. Наприклад, у роботі (Bartalesi et al., 2015) розглядається, як онтології дозволяють формалізувати знання про твори Данте Аліг'єрі, описуючи їхню стилістику, тематику та семантичні зв'язки. Цей підхід створює основу

для систематичного літературного аналізу, який може бути корисним як для дослідників, так і для освітніх цілей.

У мовознавстві спеціалізовані онтології можуть служити інструментом для моделювання граматичних структур, синтаксичних взаємозв'язків та еволюції лексичних груп. Вони дозволяють досліджувати зміни в мові, порівнювати різні мовні системи або створювати інтерактивні навчальні ресурси, такі як багатомовні словники чи граматичні довідники. Особливо корисні онтології для вивчення мов, які мають складну структуру або багату історію, оскільки вони забезпечують збереження семантичної точності та контекстуальних особливостей.

Інтеграція онтологій із сучасними технологіями, такими як великі мовні моделі, значно розширює їхні можливості. Наприклад, LLM можуть бути використані для автоматичного витягання концептів і зв'язків із текстів, що значно спрощує процес створення онтологій і дозволяє адаптувати їх до нових контекстів. Так, у роботі (Mukanova et al., 2024) було продемонстровано, як мовні моделі можуть допомогти збагачувати онтології, роблячи їх більш релевантними та функціональними, що особливо важливо, коли дослідження постійно стикаються з новими концептами, стилями й тенденціями. Не менш важливою є роль онтологій у вирішенні проблем із неоднозначністю та неточністю, які часто виникають у процесі аналізу текстів. Формалізовані моделі знань забезпечують високий рівень точності та узгодженості даних, що сприяє уникненню помилкових інтерпретацій і неправильних висновків. Вони також можуть використовуватися для виявлення та корекції помилок у результатах роботи LLM, забезпечуючи перевірку згенерованих текстів на відповідність заданим семантичним критеріям.

Розроблення спеціалізованих онтологій є складним процесом, який потребує міждисциплінарного підходу. Залучення фахівців із філології, комп'ютерних наук та інформаційних технологій дозволяє створювати моделі, які є одночасно точними, адаптивними та функціональними. Такі онтології можуть стати основою для нових цифрових платформ, які об'єднують аналіз текстів, міжмовний переклад, підсумовування текстів і автоматизовану класифікацію літературних даних. При цьому, великі мовні моделі можуть не лише використовувати онтології, а і бути використаними для їх побудови. Як показали автори у (Neuhaus, 2024), ChatGPT здатен згенерувати готову онтологію в форматі OWL (Ontology Web Language), яка в подальшому може бути розширена

та застосована для покращення якості відповідей. Важливими є також можливості онтологічного підходу для створення міждисциплінарних моделей знань. Так, у роботі (Polenghi et al., 2022) створено таку модель у галузі управління активами, яка поєднує у собі знання з різних предметних областей. Подібний підхід може бути застосований і для філологічних досліджень, наприклад, структуризації інформації про лексичні поля, синтаксичні структури чи стилістичні особливості.

Отже, спеціалізовані онтології є незамінним інструментом для проведення досліджень у галузі прикладної лінгвістики, відкриваючи нові можливості для дослідження мови та літератури. Вони сприяють систематизації знань, автоматизації рутинних процесів і створенню нових інструментів для аналізу текстів, що робить їх важливою складовою сучасних філологічних досліджень і освіти.

До викликів при створенні спеціалізованих онтологій належать, зокрема, недостатня кількість релевантних даних для побудови онтологій у вузьких галузях філології, а також складність інтеграції міждисциплінарних знань. Також актуальною проблемою є врахування специфіки мовних і культурних контекстів. Перспективними шляхами вирішення цих проблем можуть бути гнучкі методології для генерації спеціалізованих онтологій та вдосконалення інструментів для їх адаптації до конкретних завдань. У майбутньому розвиток спеціалізованих онтологій та їх інтеграція із сучасними технологіями можуть значно розширити межі гуманітарних наук, забезпечуючи якісно новий рівень роботи з текстовими даними.

Висновки. Таким чином, на основі проаналізованої літератури було досліджено потенціал інтеграції великих мовних моделей та онтологій у різні аспекти філології, включаючи семантичний та стилістичний аналіз, класифікацію текстів, генерацію текстів, пошук міжтекстових зв'язків та створення підсумків. Великі мовні моделі демонструють унікальну здатність адаптуватися до складних завдань завдяки трансформерній архітектурі, яка забезпечує контекстну обробку тексту на багатьох рівнях. Це дозволяє враховувати семантичні, синтаксичні та прагматичні аспекти мовних явищ, створюючи основи для багатогранного аналізу текстів. Використання LLM у поєднанні з онтологіями відкриває нові перспективи для філологічних досліджень, дозволяючи автоматизувати рутинні процеси, підвищити точність аналізу текстів та розширити можливості дослідників у вивченні мовних і літературних феноменів.

З'ясовано, що однією з ключових переваг великих мовних моделей є їхня здатність працювати з великими обсягами текстових даних, враховуючи контекст і багатопланові семантичні зв'язки. Водночас онтології додають структурованості та формалізації, що забезпечує більш точну інтерпретацію текстів і зменшує ризик помилкових висновків. Інтеграція цих технологій створює сприяє глибшому розумінню складних мовних явищ і відкриває нові можливості для аналізу літературних творів, стилістики, семантики та міжтекстовості. При цьому LLM стикаються з рядом викликів, зокрема, проблемою «галюцинацій» – генерації текстів, які не відповідають дійсності. Онтології можуть бути ефективним інструментом для боротьби з цими викликами, адже вони забезпечують перевірку фактологічної достовірності й узгодженість інформації. Крім того, онтології сприяють розвитку спеціалізованих моделей, які адаптовані до специфіки філологічних досліджень, наприклад, для створення міжмовних словників, аналізу стилістичних фігур чи автоматизації класифікації літературних жанрів.

Підкреслено також важливість розробки спеціалізованих онтологій для філології, які дозволяють формалізувати знання про мовні та літературні феномени, забезпечуючи систематизацію інформації і підтримку інноваційних методів аналізу. Спеціалізовані онтології надають змогу більш глибоко досліджувати міжтекстові зв'язки, стилістичні особливості та еволюцію мови, забезпечуючи точність та релевантність результатів. Завдяки своїй структурованості, вони створюють умови для об'єднання різних дисциплін, таких як мовознавство, літературознавство, та навіть культурологія, що робить їх універсальним інструментом у гуманітарних науках.

Важливим аспектом є можливість використання онтологій у взаємодії з освітніми платформами. Вони можуть допомогти у створенні динамічних навчальних матеріалів, які адаптуються до рівня знань студентів та забезпечують більш персоналізований підхід до навчання. Інтеграція онтологій із сучасними технологіями навчання може значно підвищити ефективність

освітніх процесів, сприяючи поглибленому вивченню мови, культури та літератури. Залучення міждисциплінарного підходу до створення таких онтологій може значно розширити їхні можливості.

Заслужують уваги також адаптації LLM та онтологій до багатомовного аналізу, що є особливо актуальним для досліджень у сфері порівняльного літературознавства та мовознавства. Завдяки таким технологіям стає можливим швидкий і точний аналіз великих корпусів текстів різними мовами, що сприяє виявленню культурних і лінгвістичних зв'язків. Інтеграція онтологій дозволяє автоматизувати і формалізувати процеси перекладу та адаптації текстів до різних мовних контекстів, що відкриває нові можливості для міжмовного дослідження.

Дослідження у сфері LLM та онтологій також надають змогу розробляти інструменти для роботи з історичними текстами із урахуванням їхнього стилю, семантики та культурних особливостей. Це дозволяє відтворювати контекст і мовні особливості, які є важливими для повноцінного аналізу культурної спадщини. Наприклад, створення цифрових архівів, що інтегрують онтології для структурованого представлення даних, допомагає у збереженні й популяризації історичних текстів.

Використання LLM та онтологій не лише сприяє автоматизації дослідницьких процесів, але й змінює підхід до вивчення мовних і літературних явищ. Перспективу досліджень у цій галузі вбачаємо у розвитку методів інтеграції онтологій із великими мовними моделями, вирішенні проблем із галюцинаціями та адаптації технологій до конкретних філологічних завдань. У майбутньому це забезпечить якісно новий рівень досліджень, сприятиме підвищенню ефективності освітніх процесів і відкриє нові горизонти для гуманітарних наук. Багатообіцяючим напрямком є розроблення алгоритмів, що базуються на інтеграції онтологій і LLM для виявлення прихованих закономірностей у текстах. Такі алгоритми можуть бути використані для вивчення історичних змін у мові, аналізу міжкультурних зв'язків або створення моделей для прогнозування розвитку мовних феноменів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Polenghi A., Roda I., Macchi M. Knowledge reuse for ontology modelling in Maintenance and Industrial Asset Management. *Journal of Industrial Information Integration*. 2022. Vol. 27. №. 100298. P. 1–39. DOI: 10.1016/j.jii.2021.100298.
2. Zulkipli Z. Z., Maskat R., Teo N. H. I. A systematic literature review of automatic ontology construction. *Indonesian Journal of Electrical Engineering and Computer Science*. 2022. Vol. 28. №. 2. P. 878–889. DOI: 10.11591/ijeecs.v28.i2.pp878-889.
3. Mukanova A., Milosz M., Daultekaliyeva A. LLM-powered natural language text processing for ontology enrichment. *Applied Sciences*. 2024. Vol. 14. №. 13. P. 5860–5875. DOI: 10.3390/app14135860.

4. Palagin O., Kaverinsky V., Litvin A. OntoChatGPT Information System: Ontology-Driven Structured Prompts for ChatGPT Meta-Learning. *International Journal of Computing*. 2023. Vol. 22. №. 2. P. 170-183. DOI: 10.47839/ijc.22.2.3086.
5. Yorsh U., Behr A. S., Kockmann N. Text-to-Ontology Mapping via Natural Language Processing Models. *Computers*. 2023. Vol. 12. №. 1. P. 14-39. DOI: 10.3390/computers12010014.
6. Giglou H. B., D'Souza J., Auer S. LLMs4OL: Large Language Models for Ontology Learning. *The Semantic Web – ISWC*. 2023. C. 408-427. DOI: 10.1007/978-3-031-47240-4_22.
7. Neuhaus F. Ontologies in the era of large language models – a perspective. *Applied Ontology*. 2024. Vol. 18. №. 4. P. 399-407. DOI:10.3233/AO-230072.
8. Nguyen T.H., Nguyen H. D., Pham V. T. Legal-Onto: An Ontology-based Model for Representing the Knowledge of a Legal Document. *17th International Conference on Evaluation of Novel Approaches to Software Engineering*. 2022. DOI: 10.5220/0011066300003176.
9. Toro S., Anagnostopoulos A. V., Bello S. M. Dynamic Retrieval Augmented Generation of Ontologies using Artificial Intelligence (DRAGON-AI). *Journal of Biomedical Semantics*. 2024. Vol. 15. №. 19. DOI: 10.1186/s13326-024-00320-3.
10. Dessi D., Osborne F., Recupero D. F. Generating Knowledge Graphs by Employing Natural Language Processing and Machine Learning Techniques within the Scholarly Domain. *Future Generation Computer Systems journal – Special Issue on Machine Learning and Knowledge Graphs*. 2024. Vol. 129. P. 253-264. DOI: 10.48550/arXiv.2011.01103.
11. Hu Y., Liu D., Wang Q. Automating Knowledge Discovery from Scientific Literature via LLMs: A Dual-Agent Approach with Progressive Ontology Prompting. 2024. DOI: 10.48550/arXiv.2409.00054.
12. Kowsari K., Meimandi K. J., Heidarysafa M. Text Classification Algorithms: A Survey. *Information*. 2019. Vol. 10. №. 4. P. 150-218. DOI: 10.3390/info10040150.
13. Goel A., Hira M., Anand A. Advancements in Scientific Controllable Text Generation Methods. arXiv.2307.05538. 2021. DOI:10.48550/arXiv.2307.05538.
14. Tkachenko O. I., Tkachenko K. O., Tkachenko O. A. (2021) Linguistic Ontologies: Designing and Using in the Educational Intellectual Systems. *Digital Platform Information Technologies in Sociocultural Sphere*. 2021. Vol. 4. №. 1. P. 97-111. DOI: 10.31866/2617-796X.4.1.2021.236950.
15. Vaswani A., Shazeer N., Parmar N. Attention Is All You Need. arXiv.1706.03762. 2017. DOI: 10.48550/arXiv.1706.03762.
16. Opara C. StyloAI: Distinguishing AI-Generated Content with Stylometric Analysis. *25th International Conference on Artificial Intelligence in Education*. arXiv.2405.10129. 2024. DOI: 10.48550/arXiv.2405.10129.
17. Bartalesi V., Meghini C. Using an Ontology for Representing the Knowledge on Literary Texts: the Dante Alighieri Case Study. *Semantic Web*. 2017. Vol. 8. №. 3. P. 385-394. DOI: 10.3233/SW-150198.
18. Oruganti S., Nirenburg S., English J. Automating Knowledge Acquisition for Content-Centric Cognitive Agents Using LLMs. *Proceedings of the AAAI Symposium Series*. 2023. Vol. 2. №. 1. P. 379-385. DOI: 10.1609/aaais.v2i1.27703.
19. Михайлюк А., Михайлюк О., Пилипчук О. Формування лінгвістичної онтології на базі структурованого енциклопедичного ресурсу. *Радіоелектронні і комп'ютерні системи*. 2012. №. 4. С. 81-89. URL: http://nbuv.gov.ua/UJRN/recs_2012_4_14.
20. Umphrey R., Roberts J., Roberts L. Investigating Expert-in-the-Loop LLM Discourse Patterns for Ancient Intertextual Analysis. *The 4th International Conference on Natural Language Processing for Digital Humanities (NLP4DH 2024)*. Miami, USA. Duration: Nov 16 2024. USA, 2024. P. 31-41. URL: <https://arxiv.org/abs/2409.01882>.
21. Tzirides A. O., Saini A., Zapata G. Generative AI: Implications and Applications for Education. 2023. P. 1-34. DOI: 10.48550/arXiv.2305.07605.
22. Schimmenti A., Pasqual V., Tomasi F. Structuring Authenticity Assessments on Historical Documents using LLMs. arXiv.2407.09290. 2024. URL: DOI: 10.48550/arXiv.2407.09290.
23. To H. Q., Liu M. Towards Efficient Large Language Models for Scientific Text: A Review. arXiv.2408.10729. 2024. DOI: 10.48550/arXiv.2408.10729.
24. Giglou H. B., D'Souza J., Enge F. LLMs4OM: Matching Ontologies with Large Language Models. *ESWC 2024 Special Track on LLMs for Knowledge Engineering*. Hersonissos, Greece. Duration: May 26 2024 – May 30 2024. Greece, 2024. P. 23-34. DOI: 10.13140/RG.2.2.10832.42240.
25. Nananukula N., Kejriwala M. HALO: An Ontology for Representing and Categorizing Hallucinations in Large Language Models. *SPIE Defense + Commercial Sensing (DCS 2024)*. National Harbor, Maryland, USA. Duration: April 21 2024 – April 25 2024. USA, 2024. №. 130580B. P. 1-15. DOI: 10.1117/12.3014048.
26. Agrawal G., Kumarage T., Alghamdi Z. Can Knowledge Graphs Reduce Hallucinations in LLMs Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2024 – Hybrid, Mexico City, Mexico. Duration: Jun 16 2024 → Jun 21 2024. Mexico, 2024. P. 3947-3960.
27. Sansford H., Richardson N., Maretic H. P. GraphEval: A Knowledge-Graph Based LLM Hallucination Evaluation Framework. *KiL'24: Workshop on Knowledge-infused Learning co-located with 30th ACM KDD Conference*. Barcelona, Spain. Duration: August 25 2024. Spain, 2024. P. 19-31. URL: arXiv:2407.10793v1.
28. Калініченко М. М. Застосування семантичних моделей у літературознавстві та експертизі об'єктів авторського права. *Наукові записки Національного університету «Острозька академія»: серія «Філологія»*. Острог : Вид-во НаУОА, 2023. Вип. 17. С. 59-61. DOI: 10.25264/2519-2558-2023-17(85)-59-61.

REFERENCES

1. Polenghi A., Roda I., Macchi M. (2022) Knowledge reuse for ontology modelling in Maintenance and Industrial Asset Management. *Journal of Industrial Information Integration*. 27. 100298. 1–39. DOI: 10.1016/j.jii.2021.100298.
2. Zulkupli Z. Z., Maskat R., Teo N. H. I. (2022) A systematic literature review of automatic ontology construction. *Indonesian Journal of Electrical Engineering and Computer Science*. 28 (2). 878–889. DOI: 10.11591/ijeecs.v28.i2.pp878-889.
3. Mukanova A., Milosz M., Dauletkaliyeva A. (2024) LLM-powered natural language text processing for ontology enrichment. *Applied Sciences*. 14. 13. 5860–5875. DOI:10.3390/app14135860.
4. Palagin O., Kaverinsky V., Litvin A. (2023) OntoChatGPT Information System: Ontology–Driven Structured Prompts for ChatGPT Meta–Learning. *International Journal of Computing*. 22, 2. 170–183. DOI:10.47839/ijc.22.2.3086.
5. Yorsh U., Behr A. S., Kockmann N. (2023) Text-to-Ontology Mapping via Natural Language Processing Models. *Computers*. 2023. 12, 1. 14–39. DOI: 10.3390/computers12010014.
6. Giglou H. B., D’Souza J., Auer S. (2023) LLMs4OL: Large Language Models for Ontology Learning. *The Semantic Web – ISWC*. 408–427. DOI: 10.1007/978-3-031-47240-4_22.
7. Neuhaus F. (2024) Ontologies in the era of large language models – a perspective. *Applied Ontology*. 18, 4. 399–407. DOI:10.3233/AO-230072.
8. Nguyen T.H., Nguyen H. D., Pham V. (2022) Legal-Onto: An Ontology-based Model for Representing the Knowledge of a Legal Document. *17th International Conference on Evaluation of Novel Approaches to Software Engineering*. URL: DOI: 10.5220/0011066300003176.
9. Toro S., Anagnostopoulos A. V., Bello S. M. (2024) Dynamic Retrieval Augmented Generation of Ontologies using Artificial Intelligence (DRAGON-AI). *Journal of Biomedical Semantics*. 15, 19. DOI: 10.1186/s13326-024-00320-3.
10. Dessi D., Osborne F., Recupero D. F. (2024) Generating Knowledge Graphs by Employing Natural Language Processing and Machine Learning Techniques within the Scholarly Domain. *Future Generation Computer Systems journal – Special Issue on Machine Learning and Knowledge Graphs*. 129. 253–264. DOI: 10.48550/arXiv.2011.01103.
11. Hu Y., Liu D., Wang Q. (2024) Automating Knowledge Discovery from Scientific Literature via LLMs: A Dual-Agent Approach with Progressive Ontology Prompting. DOI: 10.48550/arXiv.2409.00054.
12. Kowsari K., Meimandi K. J., Heidarysafa M. (2019) Text Classification Algorithms: A Survey. *Information*. 10, 4. 150–218. DOI: 10.3390/info10040150.
13. Goel A., Hira M., Anand A. (2021) Advancements in Scientific Controllable Text Generation Methods. arXiv.2307.05538. DOI:10.48550/arXiv.2307.05538.
14. Tkachenko O. I., Tkachenko K. O., Tkachenko O. A. (2021) Linguistic Ontologies: Designing and Using in the Educational Intellectual Systems. *Digital Platform Information Technologies in Sociocultural Sphere*. 4, 1. 97–111. DOI: 10.31866/2617-796X.4.1.2021.236950.
15. Vaswani A., Shazeer N., Parmar N. (2017) Attention Is All You Need. arXiv.1706.03762. DOI: 10.48550/arXiv.1706.03762.
16. Opara C. (2024) StyloAI: Distinguishing AI-Generated Content with Stylometric Analysis. *25th International Conference on Artificial Intelligence in Education*. arXiv.2405.10129. DOI: 10.48550/arXiv.2405.10129.
17. Bartalesi V., Meghini C. (2017) Using an Ontology for Representing the Knowledge on Literary Texts: the Dante Alighieri Case Study. *Semantic Web*. 8, 3. 385–394. DOI: 10.3233/SW-150198.
18. Oruganti S., Nirenburg S., English J. (2023) Automating Knowledge Acquisition for Content–Centric Cognitive Agents Using LLMs. *Proceedings of the AAAI Symposium Series*. 2, 1. 379–385. DOI: 10.1609/aaais.v2i1.27703.
19. Mykhailiuk A., Mykhailiuk O., Pylypchuk O. (2012) Formuvannia lingvistychnoii ontolohii na bazistrukturovanoho entsyklopedychnoho resursu [Formation of a Linguistic Ontology Based on a Structured Encyclopedic Resource]. *Radioelectric and computing systems*. 4. 81–89. URL: http://nbuv.gov.ua/UJRN/recs_2012_4_14. [In Ukrainian].
20. Umphrey R., Roberts J., Roberts L. (2024) Investigating Expert-in-the-Loop LLM Discourse Patterns for Ancient Intertextual Analysis. *The 4th International Conference on Natural Language Processing for Digital Humanities (NLP4DH 2024)*. Miami, USA. Duration: Nov 16 2024. USA, 31–41. URL: <https://arxiv.org/abs/2409.01882>.
21. Tzirides A. O., Saini A., Zapata G. (2023) Generative AI: Implications and Applications for Education. 1–34. DOI: 10.48550/arXiv.2305.07605.
22. Schimmenti A., Pasqual V., Tomasi F. (2024) Structuring Authenticity Assessments on Historical Documents using LLMs. arXiv.2407.09290. DOI: 10.48550/arXiv.2407.09290.
23. To H. Q., Liu M. (2024) Towards Efficient Large Language Models for Scientific Text: A Review. arXiv.2408.10729. DOI: 10.48550/arXiv.2408.10729.
24. Giglou H. B., D’Souza J., Enge F. (2024) LLMs4OM: Matching Ontologies with Large Language Models. *ESWC 2024 Special Track on LLMs for Knowledge Engineering*. Hersonissos, Greece. Duration: May 26 2024 – May 30 2024. Greece, 23–34. DOI: 10.13140/RG.2.2.10832.42240.
25. Nananukula N., Kejriwala M. (2024) HALO: An Ontology for Representing and Categorizing Hallucinations in Large Language Models. *SPIE Defense + Commercial Sensing (DCS 2024)*. National Harbor, Maryland, USA. Duration: April 21 2024 – April 25 2024. USA, 130580B. 1–15. DOI: 10.1117/12.3014048.
26. Agrawal G., Kumarage T., Alghamdi Z. (2024) Can Knowledge Graphs Reduce Hallucinations in LLMs Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2024 – Hybrid, Mexico City, Mexico. Duration: Jun 16 2024 → Jun 21 2024. Mexico. 3947–3960.

27. Sansford H., Richardson N., Maretic H. P. (2024) GraphEval: A Knowledge–Graph Based LLM Hallucination Evaluation Framework. *KiL'24: Workshop on Knowledge-infused Learning co-located with 30th ACM KDD Conference*. Barcelona, Spain. Duration: August 25 2024. Spain, Barcelona. 19–31. URL: arXiv:2407.10793v1.

28. Kalinichenko M. M. (2023) Zastosuvannia semantychnykh modelei u literaturoznavstvi ta ekspertyzi ob'ektiv avtorskoho prava [Application of Semantic Models in Literary Studies and Copyright Object Expertise]. *Scientific Notes of the National University «Ostroh Academy»: Series «Philology.»* Ostroh: Publishing House of Ostroh Academy National University, 17, 85. 59–61. DOI: 10.25264/2519-2558-2023-17(85)-59-61 [In Ukrainian].