

INFORMATION PLATFORM "CENTER FOR INNOVATIVE THINKING"
UKRAINIAN INSTITUTE OF SCIENTIFIC STRATEGIES
EUROPEAN UNION RESEARCH DEPARTMENT
SCIENTIFIC AND PUBLISHING CENTER "PROGRESS"

INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE

EDINBURGH
INTERNATIONAL SCIENCE FORUM

SEPTEMBER 25–27, 2026

EDINBURGH, UNITED KINGDOM

ISBN 978-617-8890-04-9

INFORMATION PLATFORM "CENTER FOR INNOVATIVE THINKING"
UKRAINIAN INSTITUTE OF SCIENTIFIC STRATEGIES
EUROPEAN UNION RESEARCH DEPARTMENT
SCIENTIFIC AND PUBLISHING CENTER "PROGRESS"

EDINBURGH INTERNATIONAL SCIENCE FORUM

PROCEEDINGS OF THE INTERNATIONAL SCIENTIFIC
AND PRACTICAL CONFERENCE

September 25-27, 2026

Edinburgh, United Kingdom

This edition was approved for publication on October 10, 2026.

Published in A4 format online on website:
https://naukainfo.com/conference?id=159

Publisher: Sole proprietor Soloviov O. V. Certificate of registration in the State Register of Publishers, Manufacturers, and Distributors of Publishing Products series DK № 8227, dated April 23, 2025.

Edinburgh, United Kingdom
2026

UDC 001.3-048.35:0/9](06)

Proceedings of the International scientific and practical conference “Edinburgh International Science Forum” (September 25-27, 2026) / Publisher website: www.naukainfo.com. - Edinburgh, United Kingdom, 2026. - 86 p.

ISBN 978-617-8890-04-9
https://doi.org/10.64828/conf-159-2026

The recommended citation for this publication is:

Shevchenko T. G. Research into the specifics of the development of performing arts in Ukraine under martial law // Edinburgh International Science Forum : proceedings of the International scientific and practical conference (September 25-27, 2026). - Edinburgh, United Kingdom : naukainfo.com, 2026. - Pp. 15-21. - URL: https://naukainfo.com/conference?id=159

Editor
Soloviov O. V.

M.Sc.Ed., M.P.A., Hon. PhD, Academic Advisor,
Head of the European Union Research Department,
Ukrainian Institute of Scientific Strategies

The collection of scientific articles is a scientific and practical publication that includes research papers by students, postgraduate students, Candidates and Doctors of Sciences, researchers, and practitioners from Ukraine, Europe, neighboring countries, and beyond. The articles reflect studies of processes and changes in the structure of modern science. This collection is intended for students, postgraduate and doctoral candidates, educators, researchers, practitioners, and all those interested in current trends in the development of modern science.

E-mail: journal@naukainfo.com

Publisher website: https://www.naukainfo.com

© Publisher website: naukainfo.com, 2026
© Ukrainian Institute of Scientific Strategies (UISS), 2026
© All authors, 2026

INFORMATION TECHNOLOGIES AND SYSTEMS

УДК 004.8

Ланде Дмитро Володимирович
доктор технічних наук, професор, завідувач кафедри інформаційної безпеки
Кузьмінський Віктор
аспірант
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
м. Київ, Україна

ЗАГАЛЬНА АРХІТЕКТУРА МЕТОДУ АВТОМАТИЗОВАНОГО ФОРМУВАННЯ ГРАФІВ ЗНАНЬ ПРО КІБЕРЗАГРОЗИ З ПОТОКІВ НОВИННИХ ДАНИХ ІЗ ЗАСТОСУВАННЯМ ТЕХНОЛОГІЇ RAG ТА ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

Анотація. Запропоновано загальну архітектуру методу перетворення новинних повідомлень на граф знань про кіберзагрози. Архітектура поєднує вилучення сутностей і відношень великими мовними моделями, пошук додаткового контексту, перевірку кандидатних тверджень та збереження їх документального походження. Визначено функціональні етапи обробки, правила обміну між компонентами й вимоги до конфігурованого середовища реалізації.

Ключові слова: кіберзагрози, OSINT, граф знань, RAG, великі мовні моделі, походження знань.

Актуальність і мета

Системи моніторингу відкритих джерел накопичують потоки повідомлень про вразливості, кіберінциденти, шкідливе програмне забезпечення та суб'єктів кіберзагроз. Відомості про один об'єкт можуть бути розподіленими між документами, мати різні назви й повторюватися у передруках. Це ускладнює встановлення зв'язків і перевірку їхнього змісту. Великі мовні моделі (Large Language Models, LLM) використовують для вилучення структурованих знань із текстів кіберрозвідки [1], а Retrieval-Augmented Generation (RAG) доповнює їхній вхідний контекст результатами пошуку в зовнішньому корпусі [2, с. 2].

Застосування LLM для вилучення концептів і зв'язків із повідомлень про кібербезпеку та семантичного індексування документів розглянуто в [3, с. 26–27]. Побудову семантичної мережі кіберзагроз на основі новинних текстів із нормалізацією вилучених сутностей описано в [4, с. 71–72]. Ці підходи створюють підґрунтя для розроблення архітектури, яка поєднує вилучення кандидатів із пошуком додаткових свідчень та перевіркою тверджень.

Метою роботи є розроблення загальної архітектури методу автоматизованого формування графів знань про кіберзагрози на основі новинних даних із застосуванням LLM та RAG. Основну увагу приділено переходу від твердження-кандидата до документально обґрунтованого зв'язку, узгодженню сутностей та забезпеченню простежуваності кожного елемента графа до першоджерела.

Подання даних і знань

Вхідними даними є послідовність новинних документів, що надходять із системи моніторингу. Для кожного документа зберігаються незмінний текст, ідентифікатор, джерело, час публікації та час фактичного надходження. На основі накопичених документів формується пошуковий корпус. Фіксована вибірка з цього потоку може використовуватися для ретроспективної апробації методу.

Результат подається у вигляді орієнтованого графа знань із відомостями про походження та уточнення тверджень:

$$G = (V, E, P, Q).$$

де V — множина нормалізованих сутностей; E — множина орієнтованих типізованих зв'язків; P — посилання на документи й текстові свідчення; Q — модальність, полярність, атрибуція та часові характеристики тверджень. Таке представлення дає змогу розрізняти фактичні й заплановані дії та зберігати контекст, необхідний для інтерпретації зв'язку.

Архітектура та послідовність обробки

Запропонована архітектура містить послідовні функціональні етапи, спільні засоби зберігання та рівень конфігурування (Рис. 1. Загальна архітектура методу). Її реалізація може використовувати спеціалізованих агентів або окремі ролі одного керованого процесу.

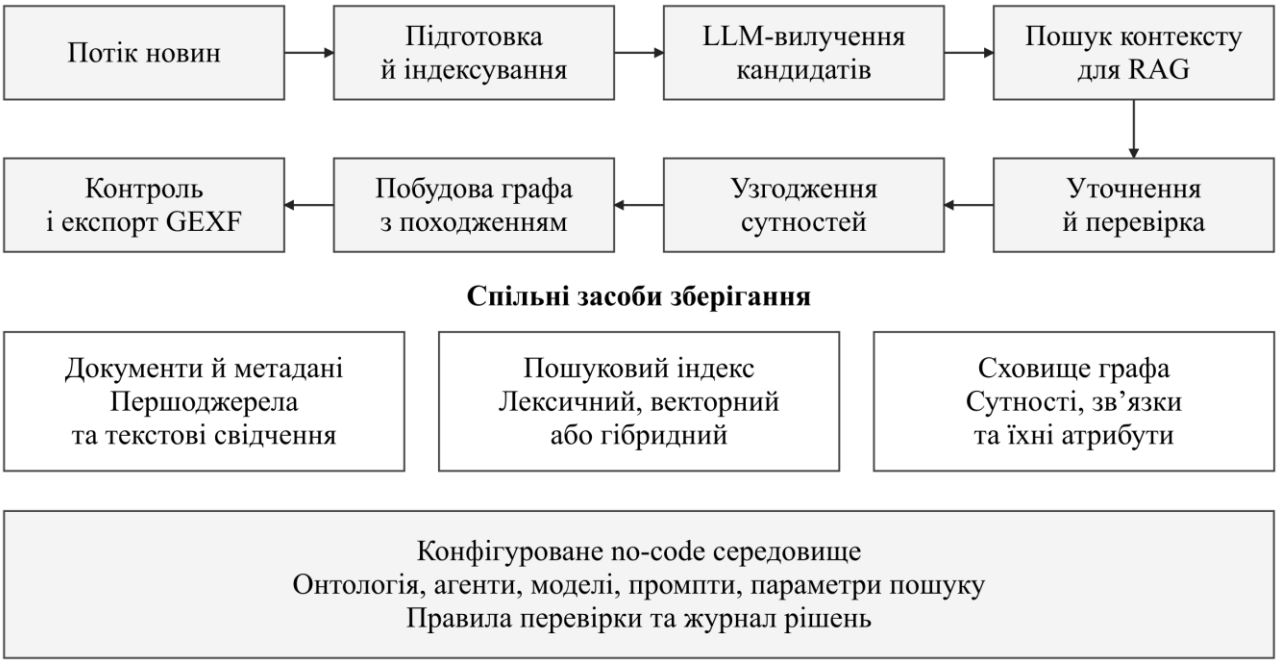


Рис. 1. Загальна архітектура методу

1. Підготовка документів. Виконуються очищення й сегментація тексту, фіксація метаданих та індексування. Ідентифікатори абзаців забезпечують адресацію доказових фрагментів. Виявлені дублікати пов'язуються з первинними записами, щоб повторна публікація не трактувалася як незалежне підтвердження.

2. Вилучення кандидатів. LLM визначає сутності та відношення-кандидати за заданою онтологією. Кандидат містить суб'єкт, предикат, об'єкт, кваліфікатори й початкове текстове свідчення. Спільна згадка двох сутностей сама по собі не є підставою для включення зв'язку до графа.

3. Пошук додаткового контексту. Запит формується з поточного повідомлення, назв сутностей або змісту твердження-кандидата. Пошуковий компонент RAG повертає релевантні фрагменти накопиченого корпусу. Він може використовувати лексичне ранжування BM25, векторну подібність або гібридний пошук; разом із фрагментами передаються посилання на джерела.

4. Уточнення й перевірка тверджень. LLM уточнює твердження-кандидат на основі первинного тексту та знайдених свідчень. Перевіряються кінцеві сутності, тип і напрям відношення, модальність та часові уточнення. Окрема перевірка відповідності результату контексту є важливою складовою RAG-систем [5, с. 1305]. Рішення має розрізняти підтримане, суперечливе та недостатньо підтверджене твердження.

5. Узгодження та побудова графа. Назви й альтернативні позначення зіставляються з урахуванням типів сутностей. Ребро включається до графа після виконання правила перевірки та збереження доказу. Суперечливі й недостатньо підтверджені кандидати залишаються в журналі. Після контролю структури результат експортується у GEXF для подальшого аналізу.

Організація програмної реалізації

Для обміну між компонентами передбачено структуровані контракти. Запис твердження-кандидата містить ідентифікатори сутностей і документа, тип відношення, текстовий фрагмент та кваліфікатори. Результат перевірки доповнює його рішенням, обґрунтуванням і посиланнями на використані джерела. Оцінка впевненості може бути додатковим атрибутом, проте не замінює документального підтвердження.

Безкодову організацію послідовної обробки текстів, формування семантичної мережі та її експорту описано в [6, с. 271–273]. Для керування ітераціями й оцінювання результатів за заданими критеріями запропоновано окремі примітиви контролю [7, с. 25–26].

Програмна архітектура розділяє сховище документів, пошуковий індекс і сховище графа. Конфігуроване no-code середовище має надавати змогу задавати онтологію, склад і послідовність агентів, мовні моделі, промпти, параметри пошуку та правила прийняття тверджень. Дослідник задає ці налаштування й виконує контрольну оцінку; повторювані операції обробки передбачається автоматизувати. Для відтворюваності зберігаються конфігурація запуску, версії промптів, результати пошуку та журнал рішень.

Для застосування до нових надходжень передбачається поповнення індексу та графа зі збереженням походження записів. За відтворення історичного стану контекст слід обмежувати документами, доступними системі на відповідний момент. Часові фільтри, контроль залежності джерел та окрема оркестрація агентів потребують самостійної перевірки; опис архітектури визначає вимоги до цих можливостей.

Висновки

Запропоновано загальну архітектуру, яка узгоджує підготовку новинних даних, вилучення кандидатних знань за допомогою LLM, пошукове доповнення, перевірку тверджень та побудову графа з документальним походженням. Визначено функціональні ролі компонентів, контракти обміну й вимоги до конфігурованого середовища. Практичне призначення архітектури полягає у формуванні структурованого подання кіберзагроз із можливістю перевірки зв’язків за джерелами. Подальше оцінювання має встановити якість вилучення, відтворюваність результатів і внесок окремих компонентів методу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Hu Y., Zou F., Han J., Sun X., Wang Y. LLM-TIKG: Threat intelligence knowledge graph construction utilizing large language model. Computers & Security. 2024. Vol. 145. Article 103999. DOI: https://doi.org/10.1016/j.cose.2024.103999
2. Lewis P. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 9459–9474. URL: https://proceedings.nips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf
3. Ланде Д. В., Рибак О. О. Семантичне індексування та кластерний аналіз документів з кібербезпеки. Реєстрація, зберігання і обробка даних. 2024. Т. 26, № 2. С. 19–32. DOI: https://doi.org/10.35681/1560-9189.2024.26.2.316711
4. Фегер А. П., Ланде Д. В. Формування семантичної мережі кіберзагроз шляхом обробки природномовних текстів. Theoretical and Applied Cybersecurity: матеріали другої Всеукраїнської науково-практичної конференції (TACS-2024). Київ: Інжиніринг, 2024. С. 71–77. URL: https://is.ipt.kpi.ua/pdf/T24.pdf
5. Sankararaman H., Yasin M. N., Sorensen T., Di Bari A., Stolcke A. Provenance: A Light-weight Fact-checker for Retrieval Augmented LLM Generation Output. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track. 2024. P. 1305–1313. DOI: https://doi.org/10.18653/v1/2024.emnlp-industry.97
6. Ланде Д. В., Рибак О. О. Безкодовий підхід до побудови семантичних мереж засобами промпт-інжинірингу. Information Technology and Security. 2025. Vol. 13, № 2. С. 264–278. DOI: https://doi.org/10.20535/2411-1031.2025.13.2.344712
7. Lande D., Danyk Y., Strashnoy L. A No-Code Programming Framework with Self-Correction Based on Large Language Models. CEUR Workshop Proceedings. 2026. Vol. 4191. P. 22–37. URL: https://ceur-ws.org/Vol-4191/paper1.pdf