МАТЕРІАЛИ
ІХ науково-практичної конференції студентів,
курсантів, аспірантів, докторантів та молодих учених
“АКТУАЛЬНІ ПИТАННЯ
ЗАСТОСУВАННЯ СПЕЦІАЛЬНИХ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМ”

з нагоди 20-ї річниці утворення
Державної служби спеціального зв’язку та захисту
інформації України

23 червня 2026 року

Київ – 2026

УДК 621

Матеріали IX науково-практичної конференції студентів, курсантів, аспірантів, докторантів та молодих учених "Актуальні питання застосування спеціальних інформаційно-комунікаційних систем". Київ : ІСЗЗІ КПІ ім. Ігоря Сікорського, 2026. 581с.

У матеріалах IX науково-практичної конференції студентів, курсантів, аспірантів, докторантів та молодих учених "Актуальні питання застосування спеціальних інформаційно-комунікаційних систем" опубліковано тези доповідей, в яких висвітлюються питання дослідження, аналізе та узагальнення нових теоретичних і практичних наукових результатів у сферах кібербезпеки, кіберзахисту, інформаційних технологій та електронних комунікацій. Підвищення професійної підготовки майбутніх офіцерів сектору безпеки і оборони шляхом формування лідерських якостей і фахових компетентностей в умовах сучасних безпекових викликів, зокрема правового режиму воєнного стану. Залучення здобувачів вищої освіти до активної наукової діяльності з використанням новітніх інформаційних технологій і штучного інтелекту в сфері кібербезпеки.

РЕЦЕНЗЕНТИ:

Олександр ПУЧКОВ	к.філос.н., професор
Сергій КОНЮШОК	к.т.н., доцент
Владислав ГОЛЬ	к.т.н., професор
Вадим РОМАНЕНКО	к.т.н., доцент
Дмитро МОГИЛЕВИЧ	д.т.н., професор
Ігор СУБАЧ	д.т.н., професор
Ярослав ЗІНЧЕНКО	к.т.н., с.н.с.

Рекомендовано до друку Вченою радою ІСЗЗІ КПІ ім. Ігоря Сікорського (протокол № 13 від 22.06.2026).

Назар ЦИНКО;
Дмитро ЛАНДЕ, д.т.н., професор

ІНТЕГРАЦІЯ ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМ ТА ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ ДЛЯ СЕМАНТИЧНОГО ПОШУКУ З ПИТАНЬ КІБЕРБЕЗПЕКИ

Анотація. У роботі проаналізовано підходи до побудови ІПС та методи семантичного пошуку з питань кібербезпеки. Розглянуто архітектуру RAG та векторний пошук, оцінено їх придатність для аналізу кіберзагроз. Розроблено концепцію інтеграції ІПС з LLM для ефективного семантичного пошуку в кібербезпеці.

Summary. The paper analyzes approaches to building information retrieval systems and semantic search in cybersecurity. The RAG architecture and vector search methods are examined and assessed for cyber threat analysis. A concept for integrating an information retrieval system with an LLM is developed.

Ключові слова: семантичний пошук, великі мовні моделі, RAG, кібербезпека, векторні представлення, OSINT, інформаційно-пошукова система.

Стрімке зростання обсягів даних про кіберзагрози – звітів про інциденти, описів CVE, публікацій OSINT – формує нові вимоги до систем інформаційного пошуку. Традиційні методи пошуку за ключовими словами (TF-IDF, BM25) орієнтовані на лексичний збіг і не здатні враховувати семантичний контекст запиту аналітика, що породжує розрив між формулюванням запиту та змістом релевантних документів. Вирішення цієї проблеми потребує підходів, що поєднують методи обробки природної мови з інтелектуальним аналізом даних про загрози.

Аналіз існуючих підходів виявив три групи методів. Класичний лексичний пошук (TF-IDF, BM25) забезпечує швидкодію, але ігнорує синонімію. Семантичний пошук на основі щільних векторних представлень (dense retrieval) кодує документи й запити bi-encoder моделями (SBERT, multilingual-E5) у спільному просторі, що дозволяє знаходити близькі за змістом документи незалежно від лексичного збігу. Гібридний пошук поєднує переваги обох підходів. Для кібербезпекового домену характерне використання баз знань MITRE ATT&CK та NVD/CVE, однак жоден підхід не забезпечує контекстуальної інтерпретації знайдених даних.

Перспективним напрямом є технологія Retrieval-Augmented Generation (RAG), яка поєднує семантичний векторний пошук з генеративними можливостями великих мовних моделей . В архітектурі RAG документи перетворюються на embeddings за допомогою encoder-моделей, зберігаються у векторному індексі (FAISS, Elasticsearch), а при запиті система витягує релевантні фрагменти та передає їх LLM для синтезу відповіді. Для кібербезпекового домену найефективнішими є спеціалізовані моделі SecBERT та CySecBERT, а також гібридний пошук з фільтрацією за метаданими. Ключова перевага RAG – оновлення бази знань без повного перенавчання, що критично для реагування на нові вразливості zero-day.

На основі проведеного аналізу сформовано вимоги та розроблено концепцію трирівневої архітектури системи: рівень збору даних (OSINT-агрегація, NER, нормалізація IoC); рівень семантичного пошуку (генерація embeddings, гібридна індексація, ранжування); рівень генерації відповідей (LLM із RAG, аналітичний звіт). Функціональні вимоги включають підтримку багатомовних запитів, фільтрацію за таксономією MITRE ATT&CK, відстеження динаміки поширення IoC та прозорість рішень через механізми XAI. Нефункціональні вимоги – обробка запитів у реальному часі та модульність архітектури.

Концепцію апробовано на даних системи CyberAggregator. Як векторне сховище обрано Elasticsearch, як LLM-компонент – Llama-2 (локальне розгортання через llama.cpp для конфіденційності даних), для візуалізації семантичних зв'язків – D3.js. Гібридний пошук підвищує точність ідентифікації релевантних документів порівняно з лексичним підходом, а інтеграція LLM забезпечує контекстуальну інтерпретацію виявлених загроз.

Висновки. Традиційні лексичні методи є недостатніми для глибокого семантичного аналізу кіберзагроз, тоді як архітектура RAG з гібридним векторним пошуком та спеціалізованими embedding-моделями суттєво підвищує релевантність результатів. Розроблена концепція трирівневої системи визначає вимоги до збору даних, семантичного індексування та генерації аналітичних відповідей із підтримкою таксономії MITRE ATT&CK та механізмів пояснювального ШІ. Система може бути ефективно використана в практичній діяльності аналітиків кібербезпеки та підрозділів SOC для автоматизованого моніторингу загроз. Подальші дослідження спрямовані на реалізацію та оцінку запропонованої архітектури на реальних масивах даних.