Олександр РИБАК;
Дмитро ЛАНДЕ, д.т.н., професор

ІНТЕЛЕКТУАЛЬНЕ РОЗШИРЕННЯ СЕМАНТИЧНИХ МЕРЕЖ НА ОСНОВІ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

Анотація. Запропоновано підхід до автоматизованого розширення семантичних мереж на основі аналізу текстів із використанням великих мовних моделей (GPT-4, Llama-3, DeepSeek-V3). Методика передбачає побудову первинної мережі понять і зв'язків, її розширення шляхом доповнення новими знаннями з LLM та інтеграцію результатів із фільтрацією слабких зв'язків.

Summary. A method is proposed for the automated expansion of semantic networks using text analysis with large language models (GPT-4, Llama-3, DeepSeek-V3). The approach includes constructing a primary network of concepts and links, expanding it using additional knowledge from LLMs, integrating results, and filtering weak links.

Ключові слова: семантичні мережі, LLM, GPT-4, Llama-3, DeepSeek-V3, OSINT, кібербезпека.

Семантичні мережі є ефективним інструментом структуризації знань у сферах кібербезпеки, OSINT та обробки природної мови, оскільки дозволяють графічно та логічно відобразити взаємозв'язки між ключовими поняттями та подіями. Вони використовуються для аналізу великих обсягів інформації, виявлення прихованих залежностей, узагальнення фактів та формування нових знань. Завдяки використанню великих мовних моделей (LLM), таких як GPT-4, Llama-3 і DeepSeek-V3, стало можливим не лише автоматично будувати такі мережі на основі вхідного тексту, але й значно їх розширювати шляхом додавання концептів і зв'язків, які не були явно представлені в джерелі, проте логічно випливають із контексту або фонових знань моделей. Це дозволяє підвищити повноту відображення знань, зменшити вплив інформаційного шуму та посилити аналітичний потенціал мережевої структури.

Запропонована методика побудови та розширення семантичних мереж включає три основні етапи:

1. Побудова початкової семантичної мережі на основі тексту за допомогою LLM, де мережа моделюється як граф

$$G = (T, E),$$

де T — множина вузлів (понять), E — множина ребер (зв'язків).

2. Розширення мережі за допомогою LLM**.** Кожна модель (GPT-4, Llama-3, DeepSeek-V3) генерує нові поняття та зв'язки на основі власної внутрішньої репрезентації знань, що дозволяє виявити додаткові, неявні семантичні зв'язки. Це розширення формально описується як

$$G' = (T \cup T', E \cup E')$$

і значно збагачує початкову структуру мережі.

3. Інтеграція результатів від кількох моделей у єдину структуру з фільтрацією слабких зв'язків. Щоб об'єднати результати декількох LLM у єдину інтегровану мережу, вводиться ваговий коефіцієнт ω_{ij} для кожного зв'язку між поняттями t_i і t_j:

$$\omega_{ij} = \frac{k}{m},$$

де k — кількість моделей, які підтвердили зв'язок, а m — загальна кількість моделей.

Щоб підвищити точність інтегрованої мережі, можна видалити малонадійні зв'язки, залишивши лише ті, для яких $\omega(a,b) > \theta$, де θ – порогове значення.

Таким чином, остаточна інтегрована мережа визначається як:

$$G_{\text{int}} = (\bigcup_m T_m, \{(t_i, t_j)\} \mid \omega_{ij} > \theta),$$

де T_m — множина термінів, згенерованих кожною з моделей.

Оцінювання моделей здійснюється за метриками точності (Precision), повноти (Recall) та F-міри для визначення якості згенерованих зв'язків. Для підвищення достовірності запропоновано фільтрацію — зберігаються лише зв'язки, підтверджені щонайменше двома моделями, що зменшує шум і забезпечує узгодженість інтегрованої мережі.

Висновки. Запропоновано технологію автоматизованого розширення семантичних мереж на основі аналізу текстів із застосуванням великих мовних моделей (LLM). Метод передбачає побудову первинної мережі понять, її збагачення новими зв'язками та поняттями, а також інтеграцію результатів із різних моделей у єдину зважену структуру. Впровадження механізму фільтрації слабких зв'язків підвищує достовірність отриманих результатів. Технологія продемонструвала ефективність у задачах текстової аналітики, зокрема в контексті кібербезпеки та OSINT.

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

18 червня 2025 року

Київ – 2025

УДК 621

Матеріали VIII науково-практичної конференції курсантів (студентів), аспірантів, докторантів та молодих учених "Актуальні питання застосування спеціальних інформаційно-комунікаційних систем". Київ : ІСЗЗІ КПІ ім. Ігоря Сікорського, 2025. 516 с.

У матеріалах VIII науково-практичної конференції курсантів (студентів), аспірантів, докторантів та молодих учених "Актуальні питання застосування спеціальних інформаційно-комунікаційних систем" опубліковано тези доповідей, в яких висвітлюються питання дослідження, аналізу й узагальнення нових теоретичних і практичних результатів у сфері кібербезпеки, новітніх інформаційних технологій, штучного інтелекту та електронних комунікацій, підготовки офіцерів для сектору безпеки та оборони а також залучення здобувачів вищої освіти до активної наукової діяльності.

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
Ірина КОНОВА	К.т.н., доцент
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 16.06.2025).

Секція № 4 "Новітні інформаційні технології та штучний інтелект в сфері кібербезпеки"