

Ланде Дмитро Володимирович

Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, м. Київ

ORCID 0000-0003-3945-1178

Пучков Олександр Олександрович

Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, м. Київ

ORCID 0000-0002-8585-1044

Субач Ігор Юрійович

Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, м. Київ

ORCID 0000-0002-9344-713X

ЗАСТОСУВАННЯ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЗНАЧЕННЯ ЦЕНТРОЇДІВ І СУТНОСТІ КЛАСТЕРІВ В МЕРЕЖІ ПОНЬЯТЬ У СФЕРІ КІБЕРБЕЗПЕКИ

Анотація. Складність інформаційного середовища вимагає нових підходів до аналізу та класифікації кіберзагроз. Одним із перспективних методів є дослідження мереж понять у кібербезпеці, виявлення ключових понять і взаємозв'язків між ключовими об'єктами та явищами у цій сфері. Ураховуючи потужний розвиток технологій штучного інтелекту цілком доцільним є використання сучасних великих лінгвістичних моделей, що можуть розглядатись в якості “віртуальних експертів”, які застосовуються під керівництвом людини-аналітика.

У статті представлено методіку визначення центроїдів та сутності кластерів у мережі об'єктів кібербезпеки на основі генеративного штучного інтелекту. Методика базується на побудові семантичної мережі понять і кластеризації мережі із застосуванням алгоритму модулярності. Процедура визначення класів модулярності здійснюється за алгоритмом, який включає кроки ініціалізації, оцінки, об'єднання кластерів, ітерації та фіналізації. Завдання визначення центроїдів понять та сутності кластерів делегується рою віртуальних експертів – генеративним мовним моделям (GPT-4, Llama-3, o1), що дозволяє автоматизувати і значно пришвидшити цей процес. Методика проілюстрована на прикладах аналізу щотижневих дайджестів з кібербезпеки. Результати показують, що застосування рою віртуальних експертів дозволяє ефективно автоматизувати процес аналізу та класифікації понять у мережі, що може стати основою для подальшого удосконалення класифікаторів кібербезпеки. Запропонована методика автоматизованого визначення центроїдів та сутностей кластерів у мережі об'єктів кібербезпеки дозволяє підвищити швидкість і об'єктивність аналізу предметної області.

Ключові слова: кібербезпека, генеративний штучний інтелект, семантичні мережі, кластеризація, центроїди, сутність кластерів.

Lande Dmytro

National technical university of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv

ORCID 0000-0003-3945-1178

Puchkov Oleksandr

National technical university of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv

ORCID 0000-0002-8585-1044

Subach Ihor

National technical university of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv

ORCID 0000-0002-9344-713X

APPLICATION OF GENERATIVE ARTIFICIAL INTELLIGENCE FOR DETERMINE THE CENTROIDS AND ESSENCE OF CLUSTERS IN THE NETWORK OF CONCEPTS IN THE FIELD OF CYBERSECURITY

Abstract. *The complexity of the information environment requires new approaches to analyzing and classifying cyber threats. One of the promising methods is the study of networks of concepts in cybersecurity, identifying key concepts and relationships between key objects and phenomena in this area. Given the strong development of artificial intelligence technologies, it is quite reasonable to use modern large linguistic models that can be considered as "virtual experts" used under the guidance of a human analyst. The article presents a methodology for determining centroids and the essence of clusters in a network of cybersecurity objects based on generative artificial intelligence. The methodology is based on building a semantic network of concepts and clustering the network using the modularity algorithm. The procedure for determining the modularity classes is carried out according to an algorithm that includes the steps of initialization, evaluation, cluster merging, iteration, and finalization. The task of determining the centroids of concepts and the essence of clusters is delegated to a swarm of virtual experts - generative language models (GPT-4, Llama-3, o1), which allows automating and significantly speeding up this process. The methodology is illustrated by analyzing weekly cybersecurity digests. The results show that the use of a swarm of virtual experts can effectively automate the process of analyzing and classifying concepts in the network, which can serve as a basis for further improvement of cybersecurity classifiers.*

The proposed method for automated determination of centroids and cluster entities in a network of cybersecurity objects allows to increase the speed and objectivity of the analysis of the subject area.

Keywords: *cybersecurity, generative artificial intelligence, semantic networks, clustering, centroids, essence of clusters.*

1. **Вступ.** Актуальність досліджень у сфері кібербезпеки постійно зростає, оскільки складність інформаційного середовища вимагає нових підходів до аналізу та класифікації кіберзагроз. Одним із перспективних методів є дослідження мереж понять у сфері кібербезпеки, що є критично важливим для розуміння динаміки загроз, виявлення ключових понять і взаємозв'язків між ключовими об'єктами та явищами у ній.

Після вирішення задачі кластеризації, тобто визначення кількості і складу кожного із кластерів, виникає завдання їх змістовної ідентифікації, найменування та вибору елементів, які найбільш точно відображають і представляють окремі кластери.

Проте, ураховуючи потужний розвиток технологій штучного інтелекту у багатьох змістовних задачах підтримки прийняття рішень, цілком доцільним є використання сучасних великих лінгвістичних моделей, що можуть розглядатись в якості "віртуальних експертів", які застосовуються під керівництвом людини-аналітика.

Великі мовні моделі (ВММ), такі як GPT-4, Llama-3 та інші, відкривають нові можливості для автоматизації завдань аналізу текстів та семантичного моделювання [1, 2]. Запропонований підхід базується на семантичному нетворкінгу [3] – методі побудови та аналізу мереж понять на основі текстових даних, а також на застосуванні рою віртуальних експертів [4] – множини запитів до генеративних моделей для отримання узгоджених результатів.

Задача побудови мереж на основі аналізу текстів у кібербезпеці вже має багато відомих рішень, як і задача кластеризації таких мереж. Зокрема, метод модулярності [5, 6], реалізований у програмному середовищі Gephi [7], дозволяє ефективно виділяти кластери у великих (складних) мережах. Проте завдання вибору центроїдів понять, надання назв кластерам та визначення їх сутності традиційно виконували експерти-люди. Передача цих функцій рою віртуальних експертів є ключовою особливістю запропонованої методики.

2. **Аналіз останніх досліджень і публікацій.** У теперішній час ГШІ є однією з найбільш перспективних галузей інформаційних технологій (ІТ), яка зосереджена на

створенні нових даних або вмісту на основі наявних даних. Системи ГШІ, такі як GPT (Generative Pre-trained Transformer 3) зайве і Gemini, вже довели свою ефективність у різних сферах, включаючи обробку природної мови, генерацію тексту, синтез мови та інші.

У рамках робіт [8, 9] було показано, як створюються і аналізуються мережі понять на основі аналізу текстів. Для цього використовуються методи екстрагування понять, виявлення зв'язків, кластеризації та алгоритми ранжирування ваги вузлів у мережі.

3. Мета і задачі дослідження. Метою проведеного дослідження була розробка методики автоматизованого визначення центроїдів та сутності кластерів у мережі об'єктів кібербезпеки за допомогою ГШІ.

Завдання статті:

- побудувати мережі понять у сфері кібербезпеки на основі аналізу текстових даних;
- кластеризувати мережу з використанням методу модулярності;
- розробити методику для визначення центроїдів та сутності кластерів з використанням рою віртуальних експертів;
- протестувати розроблену методику на прикладі аналізу щотижневих дайджестів.

4. Результати дослідження.

4.1. Формування і первинна обробка мережі понять Для формування мереж із текстів (авторами розглядалися щотижневі дайджести новин з питань кібербезпеки [10]), що включають вузли, яким відповідають поняття та зв'язки між ними, на основі застосування ВММ, застосовуються запити (промпти), наприклад (див. рис. 1):

Обери пари найбільш зв'язаних понять зі сфери кібербезпеки із наданого тексту і видай їх у вигляді стиску у форматі “поняття;поняття”. Ось текст: [текст дайджеста].

Рис. 1. Приклад промпту до системи ГШІ

При виконанні наведеного промпту отримано пари понять наступного вигляду (див. рис. 2):

*РОСІЙСЬКІ ХАКЕРИ; КІБЕРАТАКИ
РОСІЙСЬКІ СПЕЦСЛУЖБИ; ХАКЕРСЬКІ УГРУПУВАННЯ
СЛУЖБА БЕЗПЕКИ УКРАЇНИ; КІБЕРАТАКИ
КІБЕРВІЙНА; РОСІЯ
АРТ28; КРИТИЧНА ІНФРАСТРУКТУРА
ГРУ; СПЕЦСЛУЖБИ РОСІЇ
SANDWORM; INFAMOUS CHISEL
КІБЕРАТАКИ; ANDROID-ПРИСТРОЇ*

Рис. 2. Приклад відповіді системи ГШІ на промпт

При цьому, у відповідності до запропонованої концепції “рою віртуальних експертів”, сформований промпт направляється декілька разів до різних ВММ (GPT-4, Llama-3, o1). Результати, які надаються ВММ, агрегуються та фільтруються людиною-експертом, після чого зберігаються у файлі формату CSV.

4.2. Ранжування вузлів

Після отримання контенту мережі здійснюється його завантаження у систему аналізу і візуалізації графів, наприклад, Gephi [7]. Існуючі інструменти цієї системи дозволяють ранжувати отримані поняття за різними алгоритмами. Ранжування вузлів, що відповідають різним концептам здійснюється у середовищі Gephi на основі алгоритмів PageRank-TextRank [11, 12]:

$$PR(A) = \frac{1-d}{N} + d \sum_{i \in M(A)} \frac{PR(i)}{L(i)}, \quad (1)$$

де: $PR(A)$ – значення *PageRank* вузла A , d – коефіцієнт затухання, N – загальна кількість вузлів у графі, $M(A)$ – множина вузлів, що посиляються на A , $L(i)$ – кількість виходів з вузла i .

Перший член формули $\frac{1-d}{N}$ забезпечує базовий рівень *PageRank* для всіх вузлів-понять, де чисельник відповідає ймовірності такого стану, коли випадково вибирається вузол графа, а N – загальна кількість вузлів.

Другий член цієї формули $d \sum_{i \in M(A)} \frac{PR(i)}{L(i)}$ враховує центральність вузлів, які посиляються на вузол A . Для кожного вузла i , що посиляються на A (множина $M(A)$) його значення *PageRank* ділиться на кількість посилань $L(i)$, що він має. Це значення сумується для всіх вузлів, що посиляються на A .

4.3. Кластерний аналіз мережі

У середовищі програми Gephi здійснюється кластеризація сформованої мережі за класами модулярності. Існують різні види модулярності [5], які можна застосовувати в програмі Gephi [7]. У рамках запропонованої методики застосовується формула Поттса [6], яка враховує, так звану, розподільну здатність, на основі якої можна вибрати необхідну кількість класів понять (кластерів).

У методі, який визначається формулою Поттса, модифікується класичне визначення модулярності шляхом додавання параметра розподільної здатності γ , що дозволяє керувати розміром виявлених спільнот, при цьому, модулярність Q розраховується наступним чином:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta(c_i, c_j), \quad (2)$$

де A_{ij} – елемент матриці суміжності;

k_i, k_j – степені вузлів i та j ;

m – загальна кількість зв'язків у графі;

γ – параметр розподільної здатності (визначає масштаб кластеризації);

$\delta(c_i, c_j)$ – функція Кронекера.

Додавання у (2) параметра γ дозволяє адаптувати процес кластеризації до різних масштабів мережі, покращуючи точність розбиття:

– при $\gamma > 1$: формула підсилює вагу внутрішніх ребер у спільнотах, що приводить до детекції менших спільнот;

– при $\gamma < 1$: ваги внутрішніх ребер послаблюються, тому алгоритм схильний до об'єднання спільнот у більші групи;

– при $\gamma = 1$: формула збігається з класичною модулярністю Ньюмана-Гірвана [5].

Також, слід зауважити, що параметр γ дозволяє знаходити багаторівневі спільноти (*multi-resolution*), що є корисним для аналізу великих складних мереж, до яких відносяться семантичні мапи, онтології та інші моделі предметних областей.

Процедура визначення класів модулярності здійснюється за алгоритмом, який включає наступні кроки:

Ініціалізація: розподіл вузлів у початкових кластерах.

Оцінка: визначення модулярності для поточного розподілу вузлів.

Об'єднання кластерів: з'єднання вузлів або груп вузлів для визначення, чи покращиться модулярність.

Ітерація: повторення процесу об'єднання кластерів і оцінки модулярності, поки не буде досягнуто максимального значення модулярності або не буде виявлено, що подальше об'єднання не покращує результат.

Фіналізація: отримані класи модулярності представляють собою групи вузлів, що мають високу внутрішню та низьку зовнішню зв'язність.

На рис. 3 наведено кластеризований граф понять, який отримано шляхом застосування ВММ до тижневого дайджеста новин зі сфери кібербезпеки.

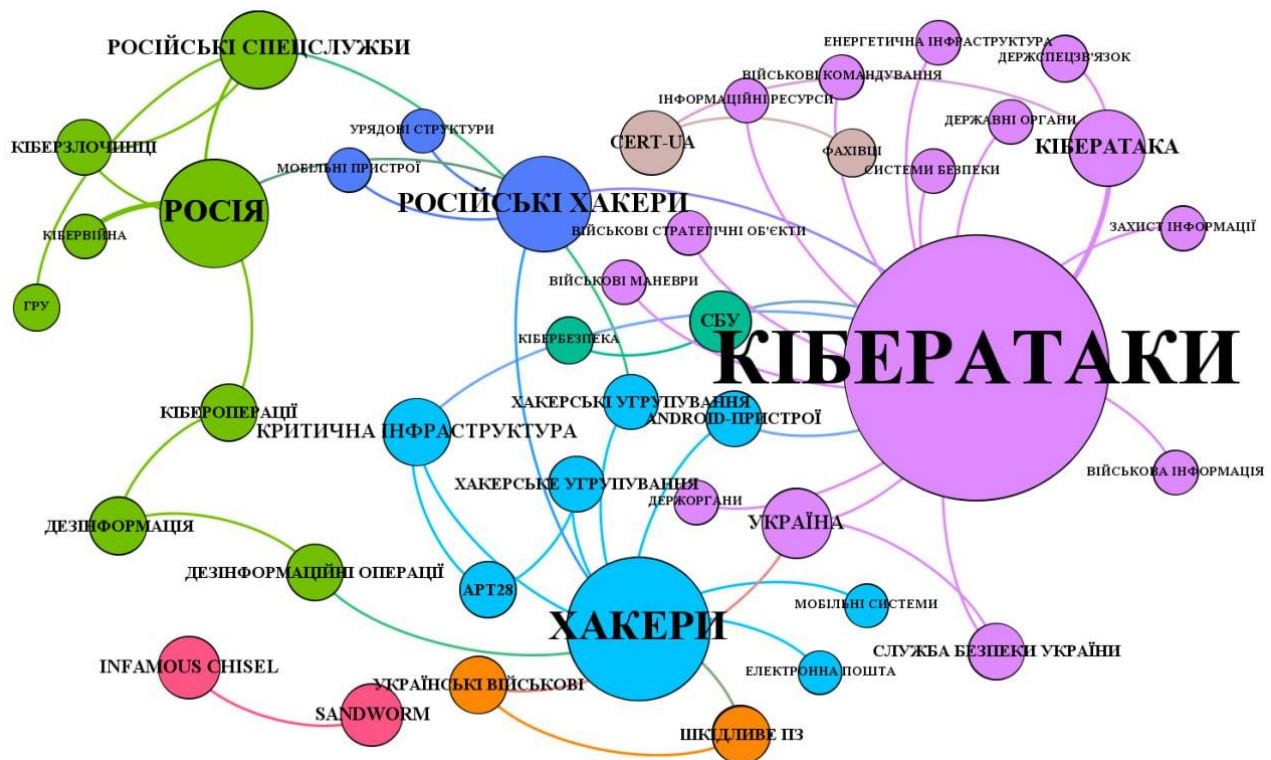


Рис. 3. Кластеризований граф понять

4.4. Визначення центроїдів і сутностей кластерів

Процедура визначення центроїду кластеру k тепер враховує не лише значення глобального *PageRank* вузлів у рамках всієї мережі, а й вагу вузлів всередині самого кластеру.

Формула для ваги вузла $w_k(v)$ у кластері k виглядає наступним чином:

$$w_k(v) = \alpha \cdot PR_{global}(v) + (1 - \alpha) \cdot PR_k(v), \quad (3)$$

де $PR_{global}(v)$ – глобальний *PageRank* вузла v у всій мережі;

$PR_k(v)$ – локальний *PageRank* вузла v у межах кластера k ;

α – ваговий коефіцієнт балансування ($0 \leq \alpha \leq 1$).

Локальний $PR_k(v)$ при обчисленні визначається для підграфа, що відповідає кластеру k за наступною формулою:

Для визначення сутності кластерів можна застосовувати промпт типу (див. рис. 6):

$$PR_k(v) = (1 - d) \cdot \frac{1}{|V_k|} + d \cdot \sum_{u \in V_k, u \rightarrow v} \frac{PR_k(u)}{\deg_k(u)}, \quad (4)$$

де d – коефіцієнт демпфування (зазвичай $d=0,85$);

V_k – кількість вузлів у кластері k ;

$u \rightarrow v$ – зв'язок між вузлом u та v у межах кластеру;

$\deg_k(u)$ – кількість вихідних зв'язків вузла u у межах кластеру.

Для визначення центроїдів і сутностей кластерів з урахуванням наведеного математичного апарату, застосовується рій віртуальних експертів, що взаємодіє з декількома генеративними моделями, наприклад, GPT-4, Llama-3 та o1 (рис. 4). В якості вхідних даних при цьому застосовується сформована мережа, перелік кластерів з вузлами та їх ваговими значеннями.

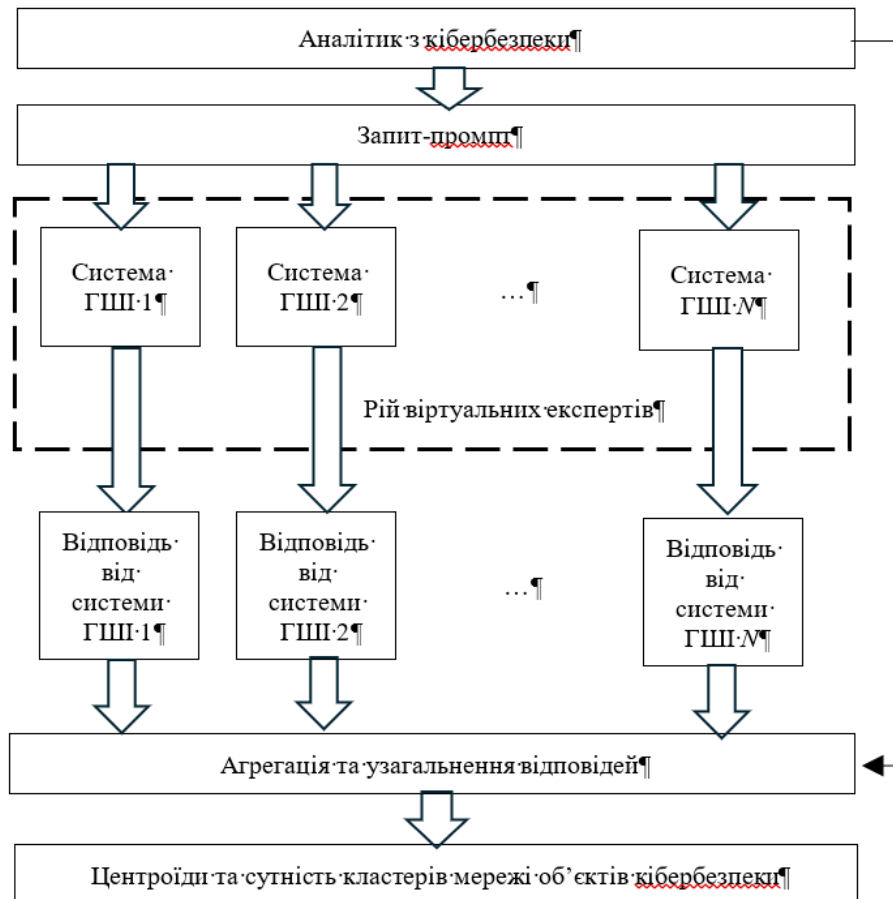


Рис.4. Схема методики визначення центроїдів і сутності кластерів мережі об'єктів кібербезпеки за допомогою систем ГШІ

Для розглянутого прикладу, визначення центроїдів кластерів застосовувався наступний промпт (див. рис. 5):

“Визначте головне поняття (центроїд) для кластера отриманої мережі, що складається з наступних вузлів: [список вузлів]”

Рис. 5. Приклад промпта для визначення центроїдів кластерів

Для визначення сутності кластерів можна застосовувати промпт типу (див. рис. 6):

“Опишіть зміст кластера для кластера отриманої мережі, що складається з наступних вузлів: [список вузлів]”

Рис. 6. Приклад промпта для визначення сутності кластерів

На практиці, обидва останніх промпта можна поєднати. Наведемо декілька прикладів промптів і відповідей на них від систем з ВММ. Для цього виберемо кластери мережі, наведеної на рис.1 та визначимо їхні центроїди і суть.

Приклад 1. Промпт до системи GPT-4 (рис. 7):

“Кластер утворюють вузли: РОСІЙСЬКІ СПЕЦСЛУЖБИ, КІБЕРВІЙНА, РОСІЯ, ГРУ, СПЕЦСЛУЖБИ РОСІЇ, ДЕЗІНФОРМАЦІЙНІ ОПЕРАЦІЇ, КІБЕРЗЛОЧИНЦІ, ДЕЗІНФОРМАЦІЯ, КІБЕРОПЕРАЦІЇ. Враховуючи PageRank всередині кластера і PageRank всередині всієї мережі, скажіть, який вузол тут центральний і кластери, що стоять поруч. Ми точно знаємо, що ці поняття - один кластер. Яке поняття є центроїдом. І одним реченням - яка суть кластера”.

Рис. 7. Промпт до системи GPT-4 (приклад 1)

Центроїдом цього кластера є “РОСІЙСЬКІ СПЕЦСЛУЖБИ”, оскільки цей вузол має найвищий PageRank як всередині кластера, так і в контексті всієї мережі. Суть кластера: цей кластер зосереджений на темах, пов'язаних з російськими спецслужбами, кіберзлочинами, кібервійною, дезінформацією та кіберопераціями, що здійснюються російськими державними структурами.

Рис. 8. Відповідь системи GPT-4 на промпт користувача (приклад 1)

Приклад 2. Промпт до системи GPT-4 (рис. 9):

“Кластер утворюють вузли: КІБЕРАТАКИ, СЛУЖБА БЕЗПЕКИ УКРАЇНИ, УКРАЇНА, ЕНЕРГЕТИЧНА ІНФРАСТРУКТУРА, ДЕРЖОРГАНИ, КІБЕРАТАКА, ВІЙСЬКОВІ МАНЕВРИ, ВІЙСЬКОВІ КОМАНДУВАННЯ, ІНФОРМАЦІЙНІ РЕСУРСИ, ВІЙСЬКОВА ІНФОРМАЦІЯ, ДЕРЖСПЕЦЗВ'ЯЗОК, СИСТЕМИ БЕЗПЕКИ, ЗАХИСТ ІНФОРМАЦІЇ, ДЕРЖАВНІ ОРГАНИ, ВІЙСЬКОВІ СТРАТЕГІЧНІ ОБ'ЄКТИ. Враховуючи PageRank всередині кластера і PageRank всередині всієї мережі, визначте, який вузол тут центральний і кластери, що стоять поруч. Ми точно знаємо, що ці поняття – один кластер. Яке поняття є центроїдом. І одним реченням – яка суть кластера”.

Рис. 9. Промпт до системи GPT-4 (приклад 2)

*Центроїдом цього кластера є "КІБЕРАТАКИ", оскільки цей вузол має найвищий PageRank як всередині кластера, так і в загальній мережі.
Суть кластера: цей кластер зосереджений на темах, пов'язаних з кіберзахистом України, включаючи державні органи, військові структури та інфраструктуру, а також забезпечення безпеки інформації та захисту критичних об'єктів від кіберзагроз.*

Рис. 10. Відповідь системи GPT-4 на промпт користувача (приклад 2)

Приклад 3. Промпт до системи GPT-4 (рис. 11):

"Кластер утворюють вузли: ХАКЕРСЬКІ УГРУПУВАННЯ, АРТ28, КРИТИЧНА ІНФРАСТРУКТУРА, ANDROID-ПРИСТРОЇ, ХАКЕРИ, МОБІЛЬНІ СИСТЕМИ, ЕЛЕКТРОННА ПОШТА. Враховуючи PageRank всередині кластера і PageRank всередині всієї мережі, визначте, який вузол тут центральний і кластери, що стоять поруч. Ми точно знаємо, що ці поняття – один кластер. Яке поняття є центроїдом. І одним реченням – яка суть кластера".

Рис. 11. Промпт до системи GPT-4 (приклад 3)

*Центроїдом цього кластера є "ХАКЕРИ", оскільки цей вузол має найвищий PageRank як всередині кластера, так і в загальній мережі.
Суть кластера: цей кластер фокусується на темах, пов'язаних із кіберзагрозами через хакерські угруповання, атаки на критичну інфраструктуру та мобільні системи, зокрема через використання Android-пристроїв та електронної пошти для атак.*

Рис. 12. Відповідь системи GPT-4 на промпт користувача (приклад 3)

Подальший процес узагальнення реалізується шляхом агрегації результатів промптів, за наступною процедурою:

Крок 1. Кожний промпт запускається до декількох систем з ВММ.

Крок 2. Результати узагальнюються з урахуванням консенсусу між відповідями.

Крок 3. Остаточне рішення приймає експерт-людина.

Висновки. У результаті дослідження розроблено методику для визначення центроїдів та сутності кластерів на основі генеративного штучного інтелекту. При цьому показано доцільність застосування рою віртуальних експертів для автоматизації задач аналізу мереж понять.

Запропонована методика автоматизованого визначення центроїдів та сутностей кластерів у мережі об'єктів кібербезпеки дозволяє підвищити швидкість і об'єктивність аналізу предметної області. Передача функцій виявлення головних понять та сутності кластерів рою віртуальних експертів є важливим кроком у напрямку автоматизації процесів вирішення задачі класифікації у кібербезпеці. Запропонована методика може стати основою для подальших досліджень у цій сфері.

Список використаної літератури

1. Jansen, B., Jung, S., Salminen, J. Employing large language models in survey research // *Natural Language Processing Journal*. 2023. Vol. 4. September. P. 100020. DOI: <https://doi.org/10.1016/j.nlp.2023.100020>.
2. Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., et al. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt // *arXiv preprint arXiv:2302.09419*, 2023.
3. Lande D., Strashnoy L. *GPT Semantic Networking: A Dream of the Semantic Web – The Time is Now*. Kyiv: Engineering, 2023. 168 p. ISBN 978-966-2344-94-3. DOI: 10.5281/zenodo.14278893.
4. Lande D., Strashnoy L. Implementation Of The Concept Of A “Swarm Of Virtual Experts” In The Formation Of Semantic Networks In The Field Of Cybersecurity Based On Large Language Models. *SSRN Preprint*. 2024. 15 p. DOI: 10.2139/ssrn.4978924.
5. Traag V. A., Waltman L., van Eck N. J. From Louvain to Leiden: guaranteeing well-connected communities // *Sci. Rep.* 2019. №9. P. 5233. DOI: 10.1038/s41598-019-41695-z.
6. Wu F. Y. The Potts model // *Rev. Mod. Phys.* 1982. №54. P. 235.
7. Bruns A., Snee H. *How to Visually Analyse Networks Using Gephi*. London: SAGE Publications, Limited, 2022. 120 p.
8. Головач Ю., Пальчиков В. Лис Микита і мережі мови // *Журнал фізичних досліджень*. 2007. Т. 11, № 1. С. 22–33.
9. Lande D., Snarskii A., Yagunova E., Pronoza E., Volskaya S. Network of Natural Terms Hierarchy as a Lightweight Ontology // *Thirteenth Mexican International Conference on Artificial Intelligence MICAI 2014*. P. 16–23. DOI: 10.1109/MICAI.2014.9.
10. Lande D., Subach I., Puchkov A. System of Analysis of Big Data from Social Media // *Information & Security: An International Journal*. 2020. Vol. 47, № 1. P. 44–61. DOI: 10.11610/isij.4703.
11. Page L., Brin S., Motwani R., Winograd T. *The PageRank Citation Ranking: Bringing Order to the Web*. Stanford InfoLab, 1999.
12. Mallick C., Das A. K., Dutta M., Sarkar A. Graph-Based Text Summarization Using Modified TextRank // *У: Nayak J., Abraham A., Krishna B., Chandra Sekhar G., Das A. (ред.) Soft Computing in Data Analytics. Advances in Intelligent Systems and Computing*. Vol. 758. Springer, Singapore, 2019. DOI: 10.1007/978-981-13-0514-6_14.