Олександр КОРИСТІН Сергій ДЕМЕДЮК

OSINT

OPEN SOURCE INTELLIGENCE

КНИГА 1

АПАРАТ РАДИ НАЦІОНАЛЬНОЇ БЕЗПЕКИ І ОБОРОНИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ

Олександр КОРИСТІН Сергій ДЕМЕДЮК

OSINT
OPEN SOURCE INTELLIGENCE

Теорія та методологія

Видано за підтримки
CRDF Global в Україні

2025

УДК 004.056.5:351.861:351.746.1:343.1

О 73

DOI: 10.32782/osint-theory-2025

Рекомендовано до друку:
вченою радою Одеського державного університету внутрішніх справ
(протокол № 11 від 26 серпня 2025 року)

Рецензенти:

Корченко О. Г. – член-кореспондент НАН України, доктор технічних наук, професор, лауреат Державної премії України в галузі науки і техніки, заслужений діяч науки і техніки України;

Корнієнко М.В. – доктор юридичних наук, професор;

Мовчан А.В. – доктор юридичних наук, професор.

OSINT Open Source Intelligence. Теорія та методологія : монографія / Користін О., Демедюк С., Барановський О., Ланде Д. та ін., за заг. ред. Користіна О.Є., Демедюка С.В. – Київ: 7БЦ, 2025. 304 с.
ISBN 978-617-8794-18-7

Монографія "OSINT: теорія та методологія" пропонує цілісну рамку осмислення розвідки з відкритих джерел як стратегічної спроможності держави та як середовища, у якому формується нова логіка реагування на виклики цифрової доби. Видання розкриває роль OSINT у національній безпеці, кіберзахисті, діяльності правоохоронних органів та у громадських ініціативах.

Методологічний блок формує нормативно-етичні та правові засади застосування OSINT: від принципів і процедур до питань територіальної юрисдикції, інтеграції інструментів у наявні інформаційні системи та автоматизації аналітичних процесів за допомогою смартконтрактів. Представлено вітчизняні інновації – семантичний нетворкінг, використання ШІ, інструменти раннього виявлення інформаційних операцій, системи ситуаційної обізнаності та аналіз цифрового сліду.

Практико-правовий вимір зосереджує увагу на застосуванні OSINT у досудовому розслідуванні: судова практика, статус і допустимість електронних доказів, використання даних із Deep Web та Dark Web і пов'язані виклики.

Монографія адресована фахівцям сектору безпеки й оборони, правоохоронцям, правникам, аналітикам, дослідникам, викладачам і здобувачам вищої освіти, які прагнуть системно опанувати теорію й методологію, позиціонуючи Україну активним суб'єктом формування практик, етичних стандартів та технологічних рішень у глобальному просторі OSINT.

УДК 004.056.5:351.861:351.746.1:343.1

ISBN 978-617-8794-18-7

РОЗДІЛ 11

СЕМАНТИЧНИЙ НЕТВОРКІНГ В ЗАДАЧАХ OSINT

Дмитро ЛАНДЕ

Семантичний нетворкінг – науково-методологічна парадигма, спрямована на автоматизоване формування, аналіз та модифікацію семантичних мереж з використанням великих мовних моделей (LLM). Основне призначення семантичного нетворкінгу – розширення можливостей аналізу даних за рахунок інтеграції методів штучного інтелекту, семантичного моделювання та обробки природної мови. Він передбачає реалізацію концепції *«рою віртуальних експертів»*, яка базується на багаторазовому опитуванні LLM для виявлення ключових понять, встановлення їхніх взаємозв'язків, оцінки значущості цих зв'язків та генерації нових знань на основі наявної інформації.

Такий підхід дозволяє не лише автоматизувати процес побудови складних семантичних структур, але й адаптувати їх до потреб аналітики в таких галузях, як OSINT, кібербезпека, стратегічне прогнозування та інтелектуальний аналіз даних.

СЕМАНТИЧНІ МЕРЕЖІ

Із семантичним нетворкінгом безпосередньо пов'язано поняття семантичних мереж, які виступають у ролі універсального інструменту моделювання та організації знань, особливо ефективно вони застосовуються в сфері Open Source Intelligence (OSINT). У контексті OSINT семантичні мережі забезпечують формалізоване представлення інформації, отриманої з різноманітних відкритих джерел, таких як веб-ресурси, соціальні мережі, наукові публікації чи новинні агентства. Вони дозволяють структурувати неструктуровані дані, встановлювати приховані зв'язки між об'єктами, а також інтегрувати розподілену інформацію в єдину аналітичну модель.

Семантичні мережі реалізуються у вигляді графів, де вершини представляють сутності (персонажі, події, географічні об'єкти тощо), а ребра — смислові або причиново-наслідкові відношення між ними. Ця властивість робить їх надзвичайно корисними для аналізу складних систем, побудови профілів суб'єктів, виявлення шаблонів поведінки та прогнозування подій на основі відкритих даних.

Перші кроки у використанні семантичних мереж було зроблено ще в 1956 році Річардом Річенсом[1] в рамках досліджень з машинного перекладу в Кембриджському центрі вивчення мов. У цій ранній концепції графова модель використовувалася для формалізації значень слів та смислових зв'язків між ними, що сприяло автоматизації лінгвістичних процесів. Цей підхід став основою для подальшого розвитку методів опрацювання природної мови, які нині широко використовуються в системах збирання, фільтрації та аналізу відкритих даних у рамках OSINT.

З початку 1980-х років концепція семантичних мереж набула нового розвитку разом із становленням веб-технологій. Особливого значення вона досягла в рамках ідеї Семантичного вебу, запропонованої Тімом Бернерсом-Лі[2] – творцем World Wide Web. Метою цієї ініціативи було створення середовища, в якому інформація була б доступною не лише для людського сприйняття, але й для автоматизованої обробки програмними засобами. Це стало важливим етапом у розвитку технологій, орієнтованих на інтеграцію, класифікацію та аналіз великих масивів відкритих даних, що має пряме відношення до сучасних

[1] Lehmann Fritz, Rodin Ervin Y., eds. (1992). Semantic networks in artificial intelligence. International series in modern applied mathematics and computer science. Vol. 24. Oxford; New York: Pergamon Press. ISBN 0080420125

[2] Berners-Lee T., Hendler J., Lassila O. The semantic web. Scientific American, 2001. Vol. 284, No. 5. – pp. 34-43.

практик OSINT.

Концепція Семантичного вебу, започаткована як наукова ідея Тіма Бернерса-Лі, передбачала створення такого інформаційного середовища, де дані були б не лише доступними для людського сприйняття, але й зрозумілими для автоматизованої обробки машинами. У цьому контексті веб як мережева структура отримав можливість автоматично інтегрувати дані з різноманітних джерел, ефективно локалізувати потрібну інформацію та формувати основу для побудови інтелектуальних систем. Семантичний веб охоплює широкий спектр стандартів, форматів та технологій, призначених для формального опису понять, термінів і взаємозв'язків між ними в межах конкретних предметних галузей. Це забезпечує структуроване представлення знань, що, в свою чергу, створює умови для глибшого аналізу, інтеграції та взаємодії інформації на глобальному рівні.

У рамках методології OSINT такі технології мають особливе значення, оскільки забезпечують формалізацію, семантичне узагальнення та аналіз масивів неструктурованих даних, зібраних із відкритих джерел — соціальних мереж, новинних порталів, наукових публікацій, офіційних документів тощо. Семантичні мережі дозволяють моделювати сутності (персонажі, події, місця, організації), встановлювати між ними причиново-наслідкові, часові, просторові чи інші типи зв'язків, що надзвичайно важливо для комплексного аналізу ситуацій, прогнозування розвитку подій, побудови профілів або виявлення шаблонів поведінки.

СЕМАНТИЧНИЙ НЕТВОРКІНГ ЯК НОВИЙ РІВЕНЬ СЕМАНТИЧНОГО МОДЕЛЮВАННЯ

Застосування класичних семантичних мереж має обмеження, пов'язані з високими витратами на їхнє проектування, наповнення та підтримку актуальності. Ці проблеми значною мірою подолані завдяки розвитку великих мовних моделей, які здійснили революцію в обробці текстової інформації. Інтеграція технологій штучного інтелекту, зокрема LLM, з семантичними мережами та концепціями Семантичного вебу, стала основою для появи нового рівня представлення знань — семантичного нетворкінгу.

Семантичний нетворкінг передбачає автоматизоване формування графів знань на основі аналізу текстових даних, що дозволяє не лише ефективно будувати семантичні мережі, але й адаптувати їх до складних, динамічно змінюваних інформаційних середовищ. Основою цієї технології є концепція *«рою віртуальних експертів»*, у якій LLM багаторазово обробляє текстові масиви, виділяючи ключові концепти та встановлюючи між ними смислові зв'язки через процедури агрегації та аналізу інформації. Такий підхід забезпечує не лише точність представлення знань, але й їхню гнучкість у подальших модифікаціях.

У процесі реалізації семантичного нетворкінгу здійснюється звернення до LLM з метою виявлення пар семантично пов'язаних понять у певній предметній області. Виявлені пари фіксуються та додаються до загальної мережі, що поступово розширюється. Це дає змогу формувати різні типи мереж: зважені та незважені, спрямовані та неспрямовані, кожна з яких може використовуватися залежно від аналітичної задачі.

Особливий науковий і практичний інтерес становлять причинно-наслідкові (каузальні) мережі, які дозволяють моделювати ланцюги взаємодії між подіями та сутностями. Такі мережі можуть слугувати основою для побудови аналітичних сценаріїв, що забезпечує глибше розуміння механізмів розвитку подій у досліджуваній області. Вони відкривають нові можливості для аналізу причинно-наслідкових зв'язків між елементами, що є критично важливим для прийняття обґрунтованих рішень на основі прогнозування та моделювання різних сценаріїв.

різними *«віртуальними експертами»*, що ускладнює побудову узгодженої мережі та вимагає додаткової фільтрації або агрегування. Крім того, якість результатів залежність від самих промптів, які відіграють ключову роль у *«рої віртуальних експертів»*, і якість відповідей безпосередньо залежить від їх точності та спрямованості. Недостатньо чітко сформульовані промпти можуть призвести до отримання нерелевантної інформації або втрати важливих зв'язків. При використанні різних моделей LLM або при запитах у різний час відповіді можуть значно відрізнятися, особливо якщо відбуваються оновлення в даних або алгоритмах моделей, що може призвести до ускладнень у підтримці послідовності результатів, що знижує достовірність і передбачуваність отримуваної інформації.

Хоча рій автоматизує значну частину роботи, він також може підтримувати творчі та аналітичні процеси людини. Віртуальні експерти можуть генерувати ідеї, які надихають на нові підходи та рішення, а також надавати людині додаткову інформацію, на основі якої можна приймати більш обґрунтовані рішення.

Перспективи розвитку *«рою віртуальних експертів»* дуже широкі, особливо з урахуванням поточного розвитку LLM та можливостей їх використання в семантичному аналізі. Майбутнє цієї методології бачиться у створенні вдосконалених механізмів агрегування та фільтрації даних для підвищення узгодженості мережі та автоматичного вирішення суперечностей між відповідями віртуальних експертів, нових моделей взаємодії між людиною та LLM, де людина зможе більш гнучко керувати роями віртуальних експертів, задаючи точні критерії, що визначають, які відповіді слід вважати значущими та достовірними; покращенні механізмів узгодження ролей і контекстів, що дозволить досягати більш точного та комплексного аналізу без істотного втручання з боку людини.

РОЗДІЛ 12

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В OSINT: ІНСТРУМЕНТИ, МЕТОДИ ТА ЕТИЧНІ ВИКЛИКИ

Олексій БАРАНОВСЬКИЙ

Сучасний світ переживає безпрецедентний інформаційний вибух. До 2025 року обсяг даних, що генеруються у світі, за прогнозами, перевищить 181 зетабайт[1], що є колосальним стрибком у порівнянні з попередніми роками. Цей феномен, що отримав назву *«великі дані»* (**Big Data**), характеризується не лише величезним обсягом (**Volume**), але й надзвичайною швидкістю (**Velocity**) надходження нової інформації та її різноманітністю (**Variety**) — від структурованих баз даних до хаотичних потоків у соціальних мережах, відео, аудіо та зображень. Для дисципліни розвідки на основі відкритих джерел, яка за своєю суттю покладається на аналіз загальнодоступної інформації, ця нова реальність стала одночасно і даром, і прокляттям.

Традиційні методи OSINT, що базувалися на ручному та напівавтоматизованому пошуку, читанні та аналізі інформації аналітиками, виявилися неспроможними впоратися з цим *«інформаційним цунамі»*. Людські когнітивні та фізичні можливості мають свої межі; один аналітик або навіть ціла команда не в змозі обробити мільйони щоденних публікацій у соціальних мережах, тисячі годин відео, що завантажуються на платформи (Youtube, TikTok, Fasebook, X тощо), чи нескінченні потоки новинних стрічок. Це призводить до низки критичних

[1] Звіт компанії Statista. https://www.statista.com/statistics/871513/worldwide-data-created/