

ефективності діяльності правоохоронних органів. Подальший розвиток аналітичних технологій, інтеграція штучного інтелекту, автоматизованої обробки великих масивів даних та міжнародного інформаційного обміну сприятимуть формуванню сучасної моделі протидії кіберзлочинності, орієнтованої не лише на розкриття, а й на попередження онлайн-шахрайства.

Список використаних джерел

1. Департамент кіберполіції Національної поліції України. Не злам, а маніпуляція: як працює соціальна інженерія у фінансовому шахрайстві. URL: <https://cyberpolice.gov.ua/article/ne-zlam-a-manipulyacziya-yak-praczuuye-soczialna-inzheneriya-u-finansovomu-shaxrajstvi-5521/>.

2. Europol. Online fraud schemes. URL: <https://www.europol.europa.eu/crime-areas/online-fraud-schemes>. Дата звернення: 17.07.2026.

3. INTERPOL. Social engineering scams. URL: <https://www.interpol.int/Crimes/Financial-crime/Social-engineering-scams>.

4. Федчак І. А. Основи кримінального аналізу : навчальний посібник. Львів : Львівський державний університет внутрішніх справ, 2021. 288 с. URL: <https://dspace.lvduvs.edu.ua/handle/1234567890/3723>.

5. Europol. Internet Organised Crime Threat Assessment (IOCTA). URL: <https://www.europol.europa.eu/publications-events/main-reports/iocta-report>.

Олейніков Олег Анатолійович,
начальник відділу програмно-технічного
забезпечення слідчої та оперативно-
розшукової діяльності Управління
інформаційних технологій Державного
бюро розслідувань

АЛГОРИТМИ ІНТЕГРОВАНОГО ПОШУКУ В НЕОДНОРІДНИХ ІНФОРМАЦІЙНИХ РЕСУРСАХ

Сучасні інформаційні системи правоохоронних органів дедалі частіше взаємодіють з неоднорідними інформаційними ресурсами, що поєднують реляційні бази даних, текстові документи, матеріали відкритих джерел, журнали подій та інші масиви. За таких умов, традиційний пошук, орієнтований на

точний або нечіткий збіг текстових фрагментів, не завжди забезпечує виявлення інформації, що описує одну й ту саму сутність, оскільки інформація може мати різні форми та повноту.

Розвиток сучасних інформаційно-пошукових систем дедалі більше пов'язується з переходом від символного пошуку до роботи зі змістовними представленнями даних. Використання векторних моделей, спеціалізованих засобів пошуку найближчих кандидатів та алгоритмів Entity Resolution – відкриває можливість інтегрувати відомості з різномірних інформаційних ресурсів, формуючи цілісне представлення об'єктів дослідження.

Традиційно алгоритми пошуку були орієнтовані на **пошук документів або окремих записів за відомими атрибутами**. Реляційні бази даних забезпечують ефективний пошук за структурованими полями, тоді як повнотекстові індекси дозволяють швидко знаходити документи за словами чи словосполученнями. Подальшим розвитком цього підходу стали алгоритми нечіткого пошуку («Fuzzy Search»), які враховують типові помилки введення, незначні відмінності у написанні та транслітерації. Такі підходи залишаються ефективними для широкого кола прикладних задач.

Одна й та сама особа може бути представлена у різних інформаційних системах як «Мельник Іван Петрович», «Мельник І.П.», «Іван Мельник», «Melnik Ivan», «Melnik I.» та мати різний набір супутніх атрибутів (державні ідентифікатори, особисті документи, дата народження, тощо). Аналогічна ситуація виникає під час збереження адрес, назв організацій, транспортних засобів та інших сутностей. Незважаючи на очевидну тотожність для людини, інформаційна система розглядає такі записи як незалежні текстові представлення.

Ускладнення задачі пошуку призводить до ще однієї важливої проблеми. Навіть алгоритми нечіткого пошуку, здатні знаходити схожі текстові представлення, порівнюють лише послідовності символів. Такий підхід дозволяє враховувати орфографічні помилки або близькі варіанти написання, проте майже не враховує зміст інформації та контекст її використання.

Подолати обмеження символного пошуку можливо завдяки **переходу до векторного представлення інформації**. Замість безпосереднього порівняння текстових рядків сучасні алгоритми дедалі частіше працюють із числовими векторами, які відображають певні властивості інформаційних об'єктів. У такому підході кожному документу, запису або окремому

текстовому фрагменту відповідає набір числових координат, що дозволяє оцінювати ступінь їх подібності за допомогою математичних методів. Це стосується як лексичної схожості (набір символів, що присутні в тексті), так і семантичної схожості (схожість за змістом слів).

Подальший розвиток алгоритмів пошуку був пов'язаний із переходом до векторного представлення інформації. Перші підходи, зокрема Bag-of-Words, TF-IDF та моделі на основі n-грам, формували вектори на основі статистичних характеристик тексту. Згодом вони були доповнені семантичними векторними представленнями (embeddings), що дозволили оцінювати змістовну близькість інформації незалежно від буквального збігу тексту. Зростання розмірності інформаційних масивів і кількості векторних представлень зумовило появу спеціалізованих векторних баз даних, які використовують алгоритми наближеного пошуку найближчих сусідів (Approximate Nearest Neighbor, ANN) для швидкого відбору найбільш релевантних результатів. Саме ці підходи сьогодні становлять технологічну основу більшості сучасних систем семантичного пошуку та Retrieval-Augmented Generation (RAG).

Однак навіть семантична близькість не завжди є достатньою підставою для висновку, що два записи описують одну й ту саму особу, організацію чи інший інформаційний об'єкт. Записи можуть бути схожими за змістом, але належати різним сутностям, або навпаки – містити мінімальну кількість спільних атрибутів, проте фактично описувати один і той самий об'єкт. За таких умов задача інтегрованого пошуку переходить від оцінювання схожості до **встановлення тотожності сутностей**, що становить предмет задачі Entity Resolution.

Для правоохоронної аналітики така ситуація є типовою. Відомості про одну особу можуть бути розподілені між кількома інформаційними системами, кожна з яких містить лише частину атрибутів. Наприклад, одна підсистема може містити прізвище, ім'я, по батькові та дату народження, інша – адресу проживання, третя – номер телефону або обліковий запис у соціальній мережі. Додаткові труднощі створюють скорочення, транслітерація, помилки введення та різні правила заповнення реквізитів.

На відміну від традиційного пошуку, результатом Entity Resolution є не перелік знайдених документів, а формування єдиного інформаційного представлення об'єкта шляхом інтеграції відомостей із різних інформаційних ресурсів.

Сучасні алгоритми Entity Resolution використовують комплексне оцінювання записів, враховуючи схожість окремих атрибутів, результати семантичного пошуку, структурні особливості даних та інші ознаки. У сучасних інформаційно-аналітичних системах векторний пошук перестає бути кінцевою метою пошукового процесу. Його основним призначенням стає формування множини найбільш імовірних кандидатів для подальшого застосування алгоритмів Entity Resolution.

Важливим джерелом додаткової інформації під час встановлення тотожності сутностей є аналіз їх взаємозв'язків. Після інтеграції даних окремі особи, організації, транспортні засоби, засоби зв'язку, фінансові операції чи події можуть бути представлені у вигляді графової структури (схеми пов'язаних сутностей). У такому представленні інформаційну цінність мають не лише власні атрибути об'єкта, а й його оточення. Наприклад, використання однакових номерів телефонів, адрес електронної пошти, транспортних засобів або регулярна участь у спільних подіях можуть суттєво підвищувати ймовірність того, що різні записи описують одну й ту саму особу.

Подальший розвиток алгоритмів Entity Resolution пов'язаний із використанням графових нейронних мереж (Graph Neural Networks, GNN), які дозволяють оцінювати схожість сутностей не лише за їх власними атрибутами, а й з урахуванням структури взаємозв'язків із сусідніми вершинами графа.

На відміну від класичних моделей, що використовують заздалегідь визначений набір ознак, **графові нейронні мережі здатні до «самонавчання»**. Перспективним напрямком розвитку інтегрованого пошуку є використання результатів роботи аналітиків як джерела накопичення навчальної вибірки. Підтверджені або відхилені рішення щодо тотожності сутностей можуть використовуватися для подальшого донавчання моделей, що дозволяє поступово підвищувати якість Entity Resolution та адаптувати алгоритми до особливостей необхідних сценаріїв пошуку.

Запропонована послідовність застосування векторних представлень, алгоритмів Entity Resolution, графових моделей та графових нейронних мереж може розглядатися як перспективний напрям розвитку інтегрованого пошуку в неоднорідних інформаційних ресурсах правоохоронних органів, забезпечуючи поступовий перехід від пошуку окремих записів до встановлення тотожності сутностей та аналізу їх взаємозв'язків.