Міністерство освіти і науки України
Кам’янець-Подільський національний університет імені Івана Огієнка
Фізико-математичний факультет
Кафедра комп’ютерних наук

Дипломна робота
магістра

з теми «ДОСЛІДЖЕННЯ ТЕХНОЛОГІЇ ДОБУВАННЯ КОРИСНОЇ ІНФОРМАЦІЇ З ВІДКРИТИХ ОНЛАЙН-ДЖЕРЕЛ ТА ЗАСОБІВ OSINT-РОЗВІДКИ»

Виконав: студент групи KN1-M22
спеціальності 122 Комп’ютерні науки
Рисюк Аспазій Вадимович

Керівник: **Смалько О.А,** доцент
кафедри комп’ютерних наук,
кандидат педагогічних наук, доцент

Рецензенти:
Оптасюк С.В, завідувач кафедри
фізики, кандидат фізико-
математичних наук, доцент;
Фурман І.Г., старший викладач
кафедри археології, спеціальних
історичних і правознавчих
дисциплін, кандидат юридичних наук

Кам’янець-Подільський – 2023

ЗМІСТ

ВСТУП

Актуальність дослідження полягає в тому, що на сьогоднішній день практично немає людини, яка б не користувалася соціальними мережами, багато злочинів, які скоюються в сучасному світі, плануються саме в інтернет-просторі, через спілкування в месенджерах, шляхом надсилання фото з зашифрованим місцем скоєння злочину тощо. OSINT-розвідка здатна ефективно протидіяти злочинам, які скоюються в кіберпросторі. Єдиною передумовою є те, що методику цю слід використовувати з урахуванням міжнародного досвіду. Одним з методів збору оперативної інформації є застосування засобів розвідувального аналізу відкритих джерел. Збільшення зацікавленості в OSINT-розвідки на сьогоднішній день спостерігається не лише з боку журналістів, аналітиків приватних компаній та пересічних громадян, але й з боку аналітиків спецслужб, оскільки ця система має певні переваги перед збором, опрацюванням та аналізом інформації з обмеженим доступом, причому насамперед у тому, що не вимагає спеціального доступу до інформації, а значить заощаджує час користувачу, а також не вимагає спеціальних навичок та істотних капіталовкладень. Використання засобів «OSINT» у деяких випадках дає змогу запобігти скоєнню злочинну, адже вислів «Краще попередити злочин, ніж за нього карати» набуває все більшого значення.

Об'єктом дослідження є сфери застосування технології OSINT-розвідки.

Предметом дослідження є методи та програмні інструменти добування корисної інформації з відкритих онлайн-джерел.

Метою кваліфікаційної роботи є дослідження засобів пошуку та аналізу інформаційних ресурсів з відкритих онлайн-джерел, вивчення можливостей їхнього оптимального застосування, а також визначення перспектив розвитку OSINT-технологій у майбутньому.

Завдання дослідження:

1. Вивчення ключових понять та принципів OSINT-технології.
2. Опис основних сфер застосування OSINT- технології.

3. Визначення технічних, програмних та інформаційних ресурсів, що використовуються у сфері OSINT.
4. Аналіз ефективності методів та інструментів для збору відкритої інформації.
5. Проектування та розробка системи, що дозволяє ефективно аналізувати та класифікувати результати пошуку відповідно до визначених критеріїв.

Методи дослідження включають в себе: вивчення наукових публікацій та літератури, що стосується добування інформації з відкритих джерел та методів OSINT-розвідки, порівняння та оцінка ефективності інструментів, використовуваних для збору та опрацювання інформації з відкритих джерел, опрацювання та аналіз отриманих в ході дослідження статистичних даних з метою визначення ефективності застосування технології, розгляд практичних прикладів застосування технології з метою добування конкретної інформації з відкритих джерел.

Робота складається з вступу, трьох розділів, списку використаних джерел, висновків і п’ятьох додатків.

Апробацію результатів дослідження здійснено у вигляді тез доповіді у збірнику матеріалів наукової конференції здобувачів вищої освіти фізико-математичного факультету Кам’янець-Подільського національного університету імені Івана Огієнка (1 листопада 2023 року) [1].

РОЗДІЛ 1.
ОСНОВИ ТЕХНОЛОГІЇ ДОБУВАННЯ ДАНИХ З ВІДКРИТИХ ДЖЕРЕЛ

1.1 Основи OSINT-розвідки

Розвідка на основі відкритих джерел (англ. Open Source INTelligence, OSINT) — концепція, методологія і технологія добування і використання військової, політичної, економічної та іншої інформації з відкритих джерел без порушення законів [3]. Вона використовується для прийняття рішень у сфері національної оборони та безпеки, в журналістиці, у сфері бізнесу, в академічному світі та наукових дослідженнях [17], [22]. що визначається як концепція пошуку військової, політичної або економічної інформації у відкритих джерелах.

Концепція OSINT передбачає збір, збереження та використання отриманої інформації в тій чи інший пошуковій системі. Зі зростанням обсягу інформації постає проблема швидкості та точності здійснення пошуку. Вирішити цю проблему дозволяє метод групування даних за певними темами. Наприклад, застосування методу семантичного пошуку покращує результативність пошуку інформації в соціальних мережах, де дані нерідко носять хаотичний характер. У міру зростання значущості інформації формується таке явище, як «інформаційна війна», в умовах якого застосування методу семантичного пошуку сприяє протидії інформаційним маніпуляціям.

Сучасний інтернет та соціальні мережі відіграють ключову роль у суспільному житті, однак можуть також бути й потенційним середовищем для планування та скоєння різноманітних злочинів. Вирішальне значення для ефективного збору та аналізу інформації з відкритих джерел в цьому контексті має застосування OSINT-технологій. Особливо важливим аспектом розвідувальної діяльності є ретельний аналіз методів, прийомів та інструментів, використовуваних у сфері застосування OSINT з метою успішного виявлення та

аналізу інформації з різноманітних відкритих джерел. Виходячи з широкого спектру застосування,

Використання OSINT допомагає державним органам спеціального призначення, військовим та розвідувальним службам на основі відкритих джерел, таких як супутникові знімки, соціальні мережі, новині ресурси тощо, забезпечувати захист державної безпеки і суверенітету країни, підтримувати конституційний лад, територіальну цілісність та оборонний потенціал, ефективно виконувати правоохоронні функції, у разі потреби планувати оборонні дії, а у випадку військових конфліктів - належно формувати військову стратегію [8], [59].

Концепція OSINT базується на таких двох основних поняттях [10]:

- відкрите джерело – це джерело інформації, яке надає її без вимоги збереження її конфіденційності, тобто інформація надається у не захищену форматі. публічного розкриття. Відкриті джерела відносяться до загальнодоступної інформації та не мають обмеження у доступі для фізичних осіб;
- загальнодоступна інформація – це інформація що опублікована та розміщена для широкого використання і є доступною для громадськості.

Вважається, що OSINT бере свій початок зі створення у Сполучених Штатах Америки в 1941 році Інформаційної служби іноземного мовлення - агентства, що використовувало розвідку з відкритим кодом в багатьох операціях Другої світової війни [9].

Види OSINT

Технічний OSINT: передбачає використання технічних інструментів і методів для збору інформації про мережі, системи та програми. Це може включати сканування портів, збір даних про доменні імена та IP-адреси, а також аналізу мережевого трафіку.

OSINT у соціальних мережах: передбачає збір і аналіз інформації з платформ соціальних мереж, таких як X, LinkedIn і Facebook. Соціальні медіа OSINT-розвідниками використовуються для виявлення потенційних загроз, моніторингу громадських настроїв і збору інформації про окремих осіб чи організації.

Dark Web OSINT: передбачає моніторинг та аналіз інформації в “темній” мережі, яка є частиною Інтернету, яка не індексується пошуковими системами та часто використовується для незаконної діяльності. Такого роду моніторинг можна використовувати для виявлення потенційних загроз, спостереження за незаконною діяльністю та збору інформації про злочинні організації [28], [56].

Фізичний OSINT: передбачає збір інформації шляхом фізичного спостереження та розвідки, як-от: відвідування сайтів, фотографування та збір інформації з публічних архівів. Фізичний OSINT можна використовувати для оцінки фізичних заходів безпеки, виявлення потенційних загроз і збору інформації про окремих осіб або організації.

Юридичний OSINT: передбачає збір і аналіз інформації з юридичних джерел, таких як протоколи судів, урядові звіти та публічні документи. Юридичний OSINT можна використовувати для збору розвідувальних даних про окремих осіб або організації, виявлення потенційних юридичних ризиків і контролю за дотриманням нормативних вимог [13].

Методи збору OSINT

Веб-пошук: використання пошукових систем, таких як Google, Yahoo, Bing тощо, для визначення інформації, пов’язаної з певною темою, зокрема даних про потенційні вразливості, нові загрози та методи атак [18].

Моніторинг соціальних медіа: моніторинг платформ соціальних медіа для виявлення потенційних загроз, моніторингу громадських настроїв і збору інформації про окремих осіб або організації.

Моніторинг темної мережі: моніторинг темної мережі для виявлення потенційних загроз, моніторингу незаконної діяльності та збору інформації про злочинні організації [20], [32].

Запити на публічні архіви: надсилання запитів на публічні архіви від державних установ та інших організацій для отримання інформації про окремих осіб або організації.

Фізичне спостереження: відвідування об’єктів та інші методи фізичного спостереження для збору інформації про заходи фізичної безпеки організації та інші вразливі місця.

Групи обміну інформацією: участь у групах обміну інформацією, наприклад в ISAC (Information Sharing and Analysis Center - Центр обміну та аналізу інформації - Міжнародна неприбуткова організація, яка вирішує глобальні проблеми кібербезпеки), а також обмін інформацією та розвідувальними даними з кваліфікованими фахівцями стосовно потенційних загроз і вразливостей.

Автоматизовані інструменти: використання автоматизованих інструментів, таких як веб-сканери та інструменти добування даних (data mining), для збору й аналізу великих обсягу інформації з різних джерел.

Ресурси OSINT-розвідки

Публічні реєстри, що надають доступ до офіційних документів, таких як судові рішення, державні звіти, тендери, патенти та інші урядові записи, мають критичне значення для OSINT. Ці джерела дозволяють дослідникам отримувати авторитетні та достовірні дані, що можуть бути використані для створення повної та точної картини ситуації або суб’єкта. Важливість цих документів полягає у їхній офіційності та загальнодоступності, що робить їх надійними джерелами для аналітичної роботи [25].

Медіа виступає важливим ресурсом для OSINT-аналітиків, які використовують традиційні засоби, такі як газети і телебачення, а також цифрові платформи, блоги, телеграм-канали та подкасти. Цифрові медіа забезпечують оперативний та гнучкий доступ до інформації, що розширює можливості в сфері OSINT [21].

Інтернет-ресурси, зокрема соціальні мережі, форуми та веб-сайти, є необхідним елементом OSINT. Соціальні мережі, такі як Facebook чи X, надають доступ до обширної інформації для відстеження трендів та збору даних про особи чи організації. Форуми та спеціалізовані веб-сайти дозволяють аналітикам заглиблюватися в конкретні теми чи спільноти. Ці джерела інформації не тільки

широко охоплюють різні аспекти, але й є динамічними та постійно оновлюються, що робить їх невід'ємними для сучасного OSINT [16].

Супутникові знімки, використовувані в OSINT, є важливим інструментом для спостереження за змінами на поверхні Землі, такими як військові рухи, міське планування та екологічні зміни. Компанії, як DigitalGlobe (частина Maxar Technologies), забезпечують високоякісні зображення для детального аналізу віддалених або важкодоступних місць. У ситуаціях, де збір інформації на місці неможливий або небезпечний, супутникові знімки надають унікальну можливість віддаленого спостереження [7].

Інформацію з відкритих джерел та публічно доступних відомостей можна отримати зі світової мережі Інтернет, дипломатичних місій, релігійних та розвідувальних організацій, академічних джерел, архівів, "сірої літератури", наукових доповідей, економічних звітів та ін. Зазначається, що інформація з відкритих джерел може мати переваги перед класифікованою таємною інформацією за критеріями оперативності, об'єму, якості, ясності, легкості використання та вартості. У випадках, коли розвідувальні дані з відкритих джерел перевершують секретну інформацію, їхню цінність обумовлено оперативністю, об'ємом, якістю, ясністю та легкістю використання. У порівнянні зі звітами розвідувальних служб, інформація з відкритих джерел відрізняється відсутністю брехні та має більшу доступність [40].

1.2 Правові підстави використання методів OSINT

У світі, де цифрова інформація стає все більш доступною, етичні та юридичні аспекти OSINT відіграють ключову роль. Розслідувачі повинні не лише бути ефективними, але й нести відповідальність за права та конфіденційність осіб, чиї дані використовуються. Стеження за юридичними рамками та етичними нормами є важливою умовою для законності та моральності розслідувань [14].

При зборі інформації з відкритих джерел необхідно утримуватися від порушення приватного життя осіб, оскільки навіть публічна інформація може включати чутливі дані. Розслідувачі повинні дотримуватися моральних норм, уникаючи надмірного втручання в особисте життя.

Знаходження балансу між необхідністю збирання інформації та захистом особистих даних є ключовим завданням. Використання інформації з соціальних мереж для розслідувань може бути корисним, проте слід дотримуватися меж конфіденційності користувачів [15].

Законодавство щодо використання OSINT може варіюватися в різних країнах. Наприклад, в Європейському Союзі регулювання конфіденційності та обробки даних є суворішим завдяки GDPR, ніж у деяких інших регіонах.

У різних країнах існують закони, які контролюють збір та використання інформації з відкритих джерел. Розслідувачі OSINT повинні бути озброєні знаннями про відповідне законодавство для уникнення порушень. Наприклад, в Україні закон "Про захист персональних даних" визначає правила обробки особистих даних .

Збір та аналіз інформації з відкритих джерел повинен відбуватися в межах нормативно-правового поля держави, забезпечуючи конституційні права та дотримуючись принципів пошуку, збору, передачі та використання інформації в усіх демократичних державах. Законодавство світових країн сприяє впровадженню систем розвідки з відкритих джерел, визначаючи обов’язки та обмеження. Наприклад, у США Закон про свободу інформації встановлює лише обмеження, пов’язані з національною обороною, фінансовою та особистою інформацією, дозволяючи вільний доступ громадянам до більшості інформації. Однак в інших країнах можуть існувати обмеження, що забороняють подібну діяльність [11].

В Україні “кожен має право вільно збирати, зберігати, використовувати і поширювати інформацію усно, письмово або в інший спосіб – на свій вибір” (Конституція України, Розділ 2, ст. 34, з урахуванням обмежень, встановлених

частиною другою статті 32). В Україні правове регулювання в інформаційній сфері ґрунтується на наступних принципах:

1) Свобода законно шукати, отримувати, передавати, виробляти та поширювати інформацію;

2) встановлювати обмеження на доступ до інформації лише законами держави;

3) відкритість інформації про діяльність державних органів та органів місцевого самоврядування та вільний доступ до такої інформації, крім випадків, передбачених законодавством держави;

4) За категорією доступу інформація поділяється на відкриту (загальнодоступну) та з обмеженим доступом.

Разом з тим, узаконеного поняття "OSINT" в Україні сьогодні не існує, хоча діяльність зі збирання, зберігання, обробки та розповсюдження інформації регулюється цілою низкою законодавчих і нормативних актів:

• Закон України "Про інформацію" від 02.10.1992 р. № 2657-XII;

• Закон України "Про охоронну діяльність" від 22.03.2012 р. № 4616-VI;

• Закон України "Про захист персональних даних" № 2297-VI від 01.06.2010 р.;

• Цивільний кодекс України (ст. 505), Кримінальний кодекс України (ст. 231, 232), Кодекс України про адміністративні правопорушення (ст. 163, ст. 163); Варто зазначити, що реалізація заходів у частині забезпечення безпеки підприємництва навіть в рамках розвідки з відкритих джерел інформації в окремих випадках сприймається як проведення оперативно-розшукової діяльності, здійснювати яку, згідно із Законом України "Про оперативно розшукову діяльність" № 2135-XII від 18.02.1992 р. вправі тільки суб'єкти, згадані в окремих статтях означених Законів України. У затвердженій Указом Президента України "Стратегії кібербезпеки України" від 15.03.2016 р. No96/2016 декларуються основні завдання силовим органам, а також передбачається "створення системи своєчасного виявлення, протидії та нейтралізації кіберзагроз, в тому числі із залученням волонтерських організацій", все це, безумовно, відноситься до застосування засобів конкурентної розвідки в цій галузі [18].

Чинний Кримінальний кодекс України передбачає кримінальну відповідальність за незаконне збирання та використання комерційної таємниці. Однак, внаслідок широкого та неоднозначного трактування законів, процедури збору, обробки та зберігання інформації про конкурентів можуть стати практично безкарними, водночас залишаючись важко доступними для громадян.

Українська практика виявляє закритий доступ до інформації, яка у інших демократичних країнах є вільно доступною, наприклад, щодо земельних ділянок, нерухомості та банківських рахунків. Більшість відомостей можна отримати лише через консультації з експертами.

Нині актуальною є проблема криміналізації державних служб, що використовують розвідку з відкритих джерел. Підрозділи служб безпеки використовують бази даних з персональними даними для позитивних цілей, але можливість незаконного доступу до таких баз, забезпечена системами типу "Cronos", створює ризик для конфіденційності особистої інформації.

Сучасні цінності визначаються конфіденційністю, правом на захист життя та свободою слова. Персональні дані стають цінним товаром, а їх недостатньою захищеністю може виникнути ризик витоку конфіденційної інформації на ринок. Захист персональних даних важливий для уникнення негативних наслідків використання цієї інформації зловмисниками [46].

Європейські стандарти захисту персональних даних включають Конвенцію Ради Європи від 28 січня 1981 року та "Пакет захисту даних" Європейського Парламенту та Ради від 27 червня 2016 року. Ці стандарти обов'язкові для країн Європейського Союзу, включаючи Україну.

Конституція України гарантує право на конфіденційність (ст. 32) та захист різних аспектів конфіденційності, включаючи територіальну, комунікаційну, інформаційну та фізичну конфіденційність (ст. 30-32, 28).

Конвенція Ради Європи визначає передачу персональних даних через кордони та визначає основні особисті дані. Законодавство України передбачає інформаційний характер обробки персональних даних, включаючи обов'язкове повідомлення та реєстрацію операторів.

Закони про захист персональних даних поширюються на всіх учасників обробки даних. Інтернет-компанії та органи розвідки, що здійснюють розвідувальну інформацію, повинні дотримуватися вимог захисту, особливо при використанні відкритих джерел.

Персональні дані, визначені Управлінням США з управління та бюджету, обговорюються як особисті і включають різні категорії, такі як ім'я, IP-адреса, ідентифікаційний номер тощо.

Захист конфіденційності персональних даних важливий для всіх учасників, і власники ресурсів повинні надавати заходи захисту та отримувати згоду суб'єктів даних.

1.3 Сфери застосування OSINT-розвідки

Розвідка в цілях безпеки – це систематичні дії та процеси, спрямовані на збір та аналіз інформації для виявлення потенційних загроз і ризиків з метою забезпечення ефективного захисту об'єкта, організації чи системи. Основні етапи розвідки включають комплексний збір інформації, глибокий аналіз ризиків, ідентифікацію конкретних загроз та розробку стратегій і заходів безпеки. Після вдосконалення систем безпеки розвідка залишається постійним процесом, який передбачає моніторинг та адаптацію стратегій та заходів до змінюючихся умов і загроз. Застосовується в різних сферах, включаючи корпоративну безпеку, національну безпеку та кібербезпеку.

Соціальна розвідка в бізнесі та конкурентна розвідка

Соціальна розвідка - стратегічний інструмент, що досліджує соціальні аспекти підприємства, включаючи споживацькі уподобання та вплив бренду через соціальні мережі. Це дозволяє адаптувати стратегії для кращого задоволення потреб аудиторії.

- Моніторинг соціальних мереж включає вивчення відгуків та активності користувачів стосовно продуктів.

- Аналіз відгуків служить для виявлення слабких місць та можливостей для покращень.
- Вивчення трендів дозволяє адаптувати продукти під зміни у споживчому попиті.
- Моніторинг активності конкурентів - ключ до формування ефективної стратегії.

Конкурентна розвідка - систематичний збір інформації про конкурентів для формування стратегічних переваг, охоплюючи аналіз продуктів, цінової політики, маркетингових стратегій, ефективності продажів та інновацій. Обидві форми розвідки є ключовими в стратегічному управлінні підприємством, надаючи інформацію для обгрунтованих рішень та досягнення конкурентних переваг.

Рекрутинг та управління кадрами Рекрутинг в галузі відкритих джерел інформації (OSINT) вимагає специфічних підходів і уваги до технічних та аналітичних навичок фахівців. Це включає аналіз потреб у розумінні областей експертизи, створення вакансій з чіткими описами, використання мережевих зв'язків для залучення кандидатів, технічний скринінг та проведення спеціалізованих інтерв'ю для оцінки аналітичних та етичних аспектів. Оцінка кібербезпеки включає перевірку навичок у використанні інструментів OSINT.

Управління кадрами для OSINT-фахівців охоплює розроблення строгих політик безпеки, надання можливостей для навчання та розвитку, забезпечення високого рівня безпеки інформації, створення командної роботи, встановлення етичних стандартів та розроблення систем винагород і стимулювання для залучення талановитих фахівців у галузі OSINT.

HR-інжиніринг для OSINT-фахівців включає спеціальні підходи до привертання, відбору та управління персоналом з урахуванням особливостей галузі відкритих джерел інформації. Основні етапи цього процесу включають аналіз області OSINT для визначення необхідних навичок, створення детальних описів вакансій з урахуванням технічних вимог, використання мережевих ресурсів для залучення кандидатів, проведення технічного скринінгу та спеціалізованих інтерв'ю для оцінки аналітичних та етичних аспектів. Оцінка рівня кібербезпеки кандидатів, надання можливостей для навчання та розвитку, а також розроблення систем винагород і стимулювання допомагають створювати

ефективні команди для роботи в умовах високотехнологічного та динамічного середовища. [19].

Розслідування правопорушень – це використання відкритих джерел інформації для збору та аналізу даних, пов'язаних із можливими або фактичними порушеннями закону. Така діяльність включає в себе вивчення відкритих ЗМІ та новин, публічних записів, офіційних заяв та звітів правоохоронних органів, судових інстанцій, урядових установ та інших відомств. Також при цьому необхідно проводити пошук і аналіз інформації з публічних базах даних, що можуть містити дані про осіб, конкретні події та компанії, діяльність яких потребує вивчення; здійснювати моніторинг соціальних мереж, блогів, форумів та інших неформальних джерел інформації для виявлення публічної активності, пов'язаної з обставини правопорушень; аналізувати взаємодії окремих осіб і груп людей в контексті предмету розслідування, у тому числі досліджуючи вміст електронної пошти та інші форми взаємодії комунікантів для виявлення можливих планів та/або їхніх дій, що порушують закон; вивчати інформацію в архівах та історичних джерелах, що можуть містити корисний контекст або попередні факти, світлини та відеоматеріали для виявлення можливих доказів чи деталей правопорушень.

Технічні аспекти інформаційної безпеки передбачають діяльність, спрямовану на вивчення і аналіз технічних складових систем, мереж, програмного забезпечення та інфраструктури з метою забезпечення та підвищення рівня безпеки. Основний акцент робиться на інформації, яка може бути зібрана з відкритих джерел, для розуміння потенційних загроз та вразливостей. Зокрема, проводиться аналіз кіберзагроз (OSINT-розвідники аналізують відкриті джерела для виявлення і вивчення кіберзагроз, таких як нові види вірусів, шкідливих програм, визначають застосовані методи атак і наявні вразливості) [55]; моніторинг вразливості безпеки і мережевого трафіку (слідкування за вчасним виправленням вразливостей в системах, різних програмних продуктах, вебзастосунках, вебсайтах та інших інтернет-ресурсах); спостереження за системами керування ідентифікацією та доступом для виявлення можливих

слабких місць і оцінювання наявного рівня безпеки; моніторинг мережевою активністю та аналіз трафіку для виявлення аномалій і потенційно шкідливої діяльності; аналіз архітектури систем різного типу (у тому числі хмарних сервісів, систем Інтернету речей [IoT]) для виявлення можливих слабких місць та точок доступу для потенційних атак; вивчення публічної інформації про компанії або організації, їхньої технічної інфраструктури, впроваджену безпекову політику, передбачувані заходи реагування на інциденти; спостереження за обговореннями актуальних питань інформаційної безпеки в професійних колах, за обміном думками в інтернет-спільнотах стосовно досліджень вразливостей; моніторинг діяльності зловмисників та хакерських груп для попередження можливих атак.

Вивчення діяльності громадських організацій, суспільної думки та громадянських ініціатив – це діяльність OSINT-розвідників, спрямована на: збирання і аналіз відкритої інформації для отримання детального бачення впливу громадських організацій та їхньої діяльності на суспільство (збір інформації про мету, цілі, стратегії та результати роботи громадських організацій; визначення ключових фігур, лідерів та активістів, що впливають на діяльність організацій); оцінювання реакцій людей на діяльність громадських ініціатив та організацій (аналіз взаємодії зацікавлених сторін, рівня громадської підтримки); вивчення прозорості діяльності та фінансування громадських організацій (дослідження відкритості та прозорості фінансової діяльності громадських ініціатив, з'ясування джерел фінансування, можливих впливових осіб та/або організацій); аналіз співпраці та партнерства (вивчення мереж співробітництва та партнерства громадських організацій, дослідження взаємодії з урядовими структурами, різними неприбутковими організаціями, бізнесом та іншими стейкхолдерами); моніторинг суспільної думки (аналіз публічних висловлювань, коментарів та вражень стосовно громадських ініціатив у соціальних мережах та ЗМІ, виявлення тенденцій та змін у суспільній думці); визначення стратегій та цілей (вивчення плану дій та мети громадських ініціатив, аналіз досягнень та планування можливих шляхів удосконалення діяльності); оцінювання впливу на суспільство

(визначення впливу громадських організацій на різні аспекти суспільства, аналіз результатів та ефективності реалізації ініціатив) [25].

Висновки до розділу

Результати дослідження підтверджують, що OSINT є ключовою складовою розвідки, що за рахунок публічної інформації забезпечує безпеку та проведення ефективних розвідувальних заходів. Основні методи та інструменти OSINT призначені для отримання об'єктивних та релевантних даних. Вивчення ключових елементів з використанням публічної інформації є стратегічно важливою частиною розвідувального процесу, що забезпечує глибоке розуміння ситуації та прийняття обґрунтованих рішень. Одночасно слід підкреслити важливість дотримання етичних та правових норм під час застосування методів OSINT для забезпечення легітимності діяльності. Розширення сфери застосування OSINT-розвідки ґрунтується на її універсальності в умовах аналізу кібербезпеки та вирішення геополітичних конфліктів, яка гарантує високий рівень інформаційної обізнаності.

РОЗДІЛ 2.
ДОСЛІДЖЕННЯ ТЕХНОЛОГІЇ ДОБУВАННЯ КОРИСНИХ ДАНИХ З ВІДКРИТИХ ДЖЕРЕЛ

2.1. Методології роботи OSINT-розвідників

Існує кілька відомих методологій роботи OSINT-розвіднкиів, які мають чітко визначений життєвий цикл. Вони використовують для збору та аналізу інформації з відкритих джерел, охоплюють планування, обробку, звітування та подальші дії.

Методологія "Оцінка загроз та можливостей" (Threat and Opportunity Assessment - TOA) у контексті OSINT-розвідки спрямована на комплексну оцінку загроз та можливостей через систематичний збір та аналіз інформації з відкритих джерел (рисунок 2.1). Основна мета TOA полягає в забезпеченні розуміння потенційних ризиків та переваг для обґрунтованих рішень. Процес TOA включає ідентифікацію цілей, оцінку загроз, визначення можливостей, аналіз ризиків та вигідності, розробку [37].

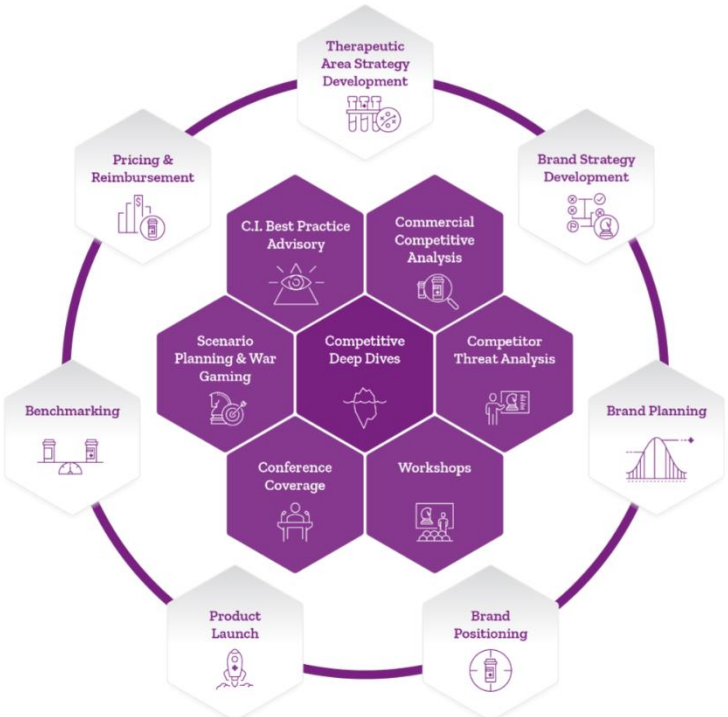

Рисунок 2.1 Систематичний збір та аналіз інформації з відкритих джерел

Методологія "Коло OSINT" (OSINT Cycle) — це систематичний та циклічний підхід до відкритої розвідки, який включає етапи підготовки, збору, обробки, аналізу та розповсюдження інформації з відкритих джерел. Основна мета полягає в створенні повного обсягу інформації та забезпеченні актуальності даних. Підготовка включає визначення цілей, потреб та вибір оптимальних джерел інформації. Збір передбачає отримання необроблених даних з різних джерел за допомогою ручних та автоматичних методів. Обробка включає перевірку автентичності та фільтрацію отриманих даних перед подальшим впорядкуванням та зберіганням (рисунок 2.2). Аналіз полягає в ретельному вивченні даних для виявлення закономірностей та ідей, з використанням візуалізації та автоматизованих інструментів. Завершенням циклу є розповсюдження розвідувального звіту серед зацікавлених осіб та організацій [38].

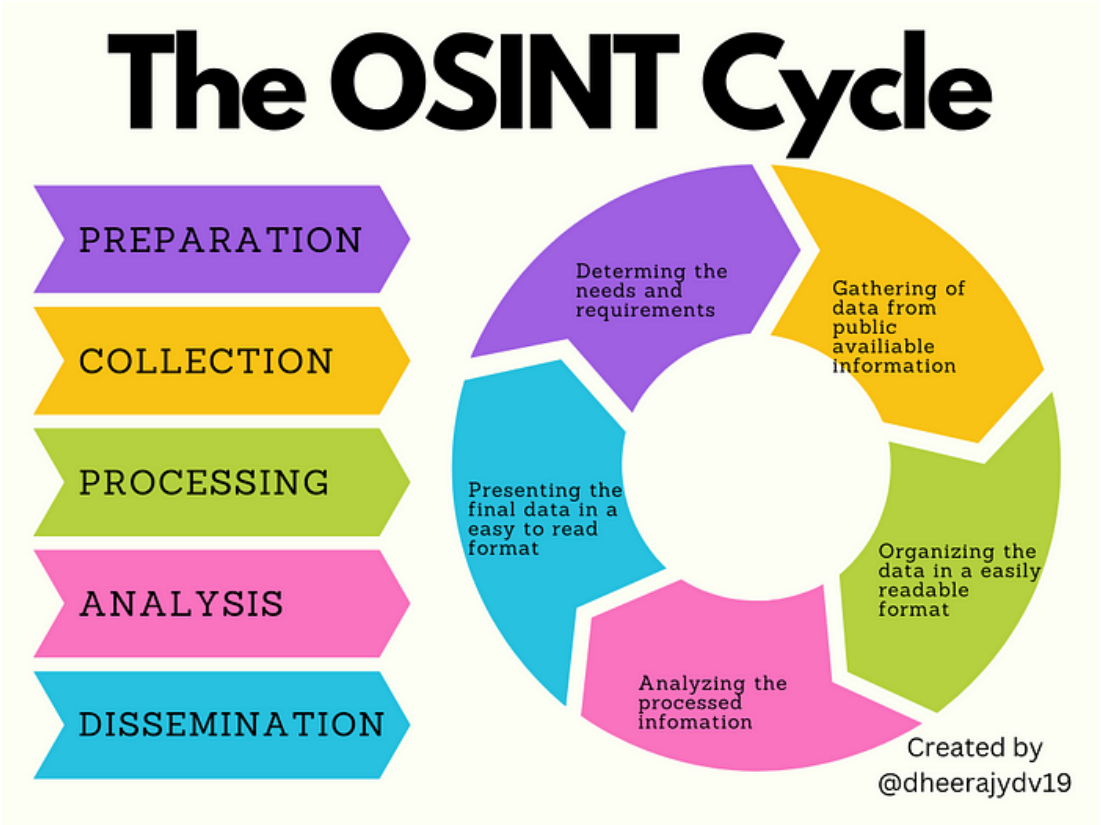

Рисунок 2.2 Схема циклічної обробки інформації

Методологія "Дослідження цифрових слідів" (Digital Footprint Investigation) спрямована на збір та аналіз цифрового сліду особи або організації в Інтернеті. Вона включає пошук основної інформації за ім'ям та прізвищем, електронною адресою та номерами телефонів. Також проводиться аналіз соціальних мереж, включаючи профілі, публікації, лайки та коментарі(рисунок 2.3). Дослідження охоплює аналіз доменів, історії сайтів, статей та блогів, а також перевірку результатів пошукових запитів та видаленої інформації. Застосовуються засоби пошуку зображень та відео, аналіз географічних даних та використання спеціалізованих інструментів для OSINT-розвідки, таких як Maltego [2], Creepy, SpiderFoot тощо. Ця методологія дозволяє створити повний образ цифрового сліду об'єкта дослідження, що є ключовим для кібербезпеки, правоохоронних дій, аналізу ризиків та інших галузей [39].

Рисунок 2.3 Схематичне зображення методології Дослідження Цифрових Слідів збір та аналіз цифрового сліду

2.2. Програмні OSINT-інструменти

OSINT, або розвідка на основі відкритих джерел, використовує програмні інструменти та методології для аналізу, збору інформації gji. Використання штучного інтелекту в цьому процесі необхідно для автоматизації збору даних та підвищення точності аналізу. Розвиток технології OSINT включає в себе використання штучного інтелекту та машинного навчання (AI/ML) для підтримки досліджень [26]. Урядові та розвідувальні служби вже використовують штучний інтелект для збору та аналізу соціальних мереж, а військові використовують AI/ML для боротьби з тероризмом та кіберзлочинністю. У приватному секторі технології AI/ML можуть поліпшити збір даних, фільтрацію шуму, аналіз і співвіднесення інформації, що допомагає аналізувати більше необроблених даних та отримувати дієві ідеї [50], [63].

Maltego, потужний інструмент для аналізу та візуалізації інформації в галузі відкритої розвідки (OSINT), збирає, аналізує та візуалізує дані з різних джерел, надаючи повну картину особи, організації чи теми (рисунок 2.4). Графічний інтерфейс спрощує взаємодію з графами даних, а засоби збору даних інтегруються з соціальними мережами, WHOIS, географічними сервісами. Автоматизовані запити дозволяють отримувати інформацію швидко, а інтерактивна взаємодія з результатами та аналіз зв'язків роблять Maltego ефективним і гнучким інструментом для OSINT [15].

Недоліки Maltego включають вартість ліцензії та обмежену функціональність у безкоштовній версії. Залежність від доступу до джерел даних та вивчення кривої можуть стати викликом для початківців. Конфіденційність та приватність повинні бути враховані при використанні для аналізу особистої інформації. Застосовуючи Maltego та інші інструменти OSINT, важливо дотримуватися законів та етичних норм, враховуючи умови користування та відмову від відповідальності розробників [24].

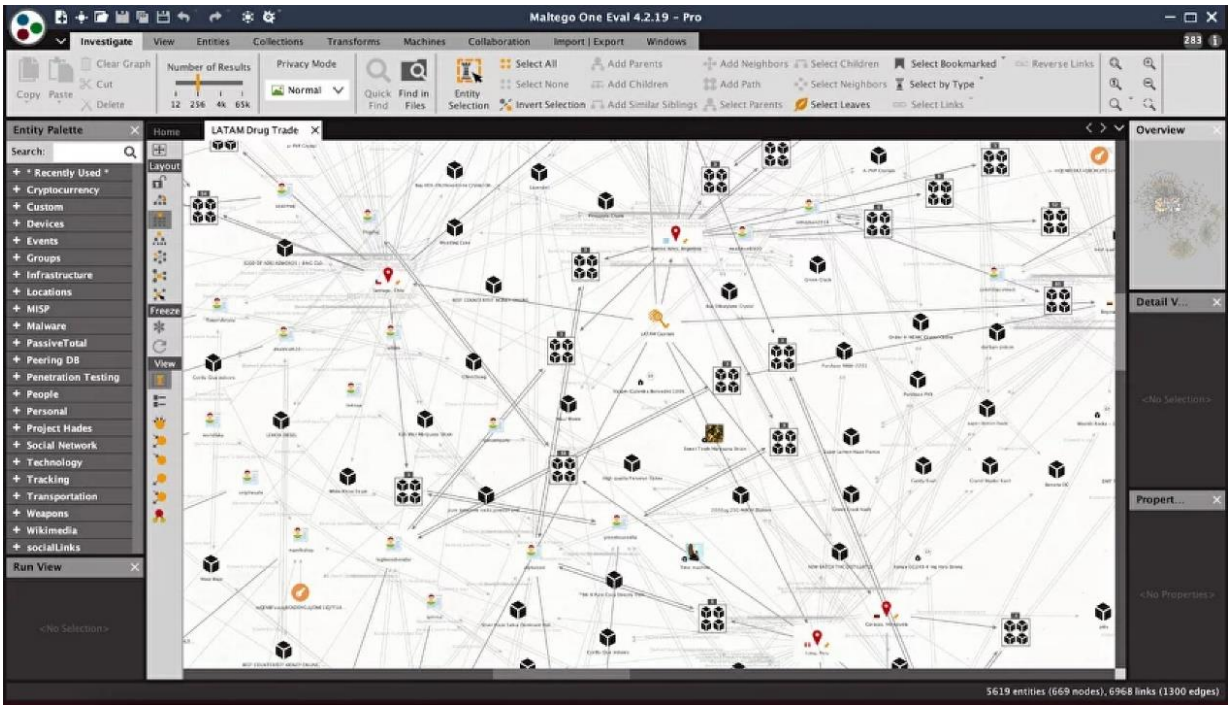

Рисунок 2.4 Візуалізація збору даних з різних джерел

Shodan — це сервіс для пошуку та аналізу пристроїв, підключених до Інтернету, таких як камери, сервери та інші IoT-пристрої (рисунок 2.5). У сфері OSINT він використовується розслідувачами для виявлення вразливих об'єктів, зокрема для ідентифікації пристроїв, які можуть стати об'єктом кібератак чи моніторингу. Зокрема, він дозволяє здійснювати пошук за технічними характеристиками, географічним розташуванням, виявляти вразливості та використовувати API для автоматизації та інтеграції з іншими інструментами. Однак важливо використовувати Shodan відповідно до законів та етичних норм, дотримуючись принципів конфіденційності, уникаючи порушення приватності та дотримуючись умов користування [42].

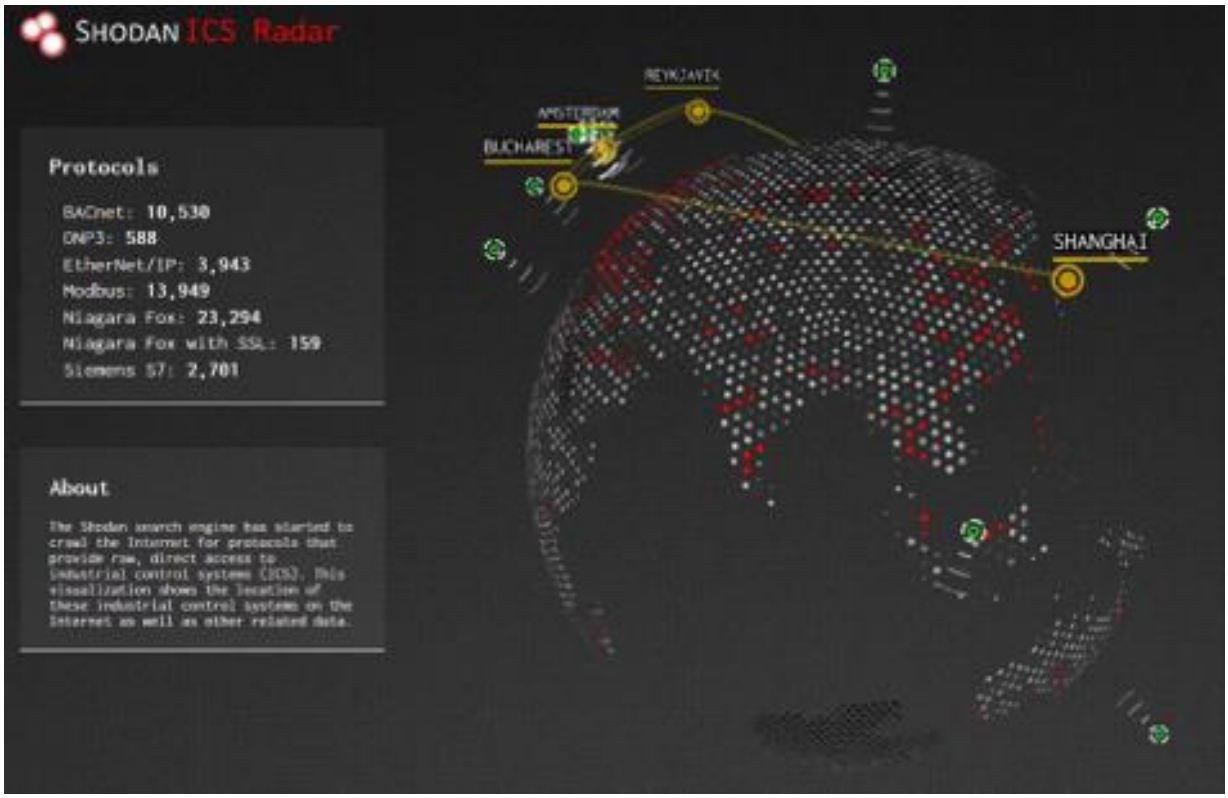

Рисунок 2.5 Пошук та аналіз пристроїв

У сфері OSINT, **Google Dorks** використовуються для швидкого виявлення конфіденційних документів та інших даних. Вони базуються на спеціальних пошукових запитах у Google, що дозволяють отримати точні та специфічні результати. Google Dorks можуть бути налаштовані для пошуку різноманітної інформації, використовуючи ключові слова, оператори та комбінації запитів. Вони також дозволяють визначати область пошуку, використовуючи параметри та оператори, такі як "intitle:" чи "filetype:". Основна вага на ефективному використанні Google Dorks в OSINT для знаходження конкретної та корисної інформації при вивченні різних аспектів безпеки та конфіденційності (рисунок 2.6). Треба враховувати їх обмеження та можливі ризики, дотримуючись етичних та правових норм використання, а також забезпечуючи захист приватності та прав осіб, що можуть бути задіяні в процесі дослідження [43].

Приклади Google Dorks:

Пошук витоків даних:

filetype: sql intext : username password

Пошук веб-камер:

inurl : "view/index.shtml "

Пошук вразливостей веб-сайтів:

intitle: "Index of" config. php

Пошук конфіденційної інформації:

site : example.com confidential filetype : pdf

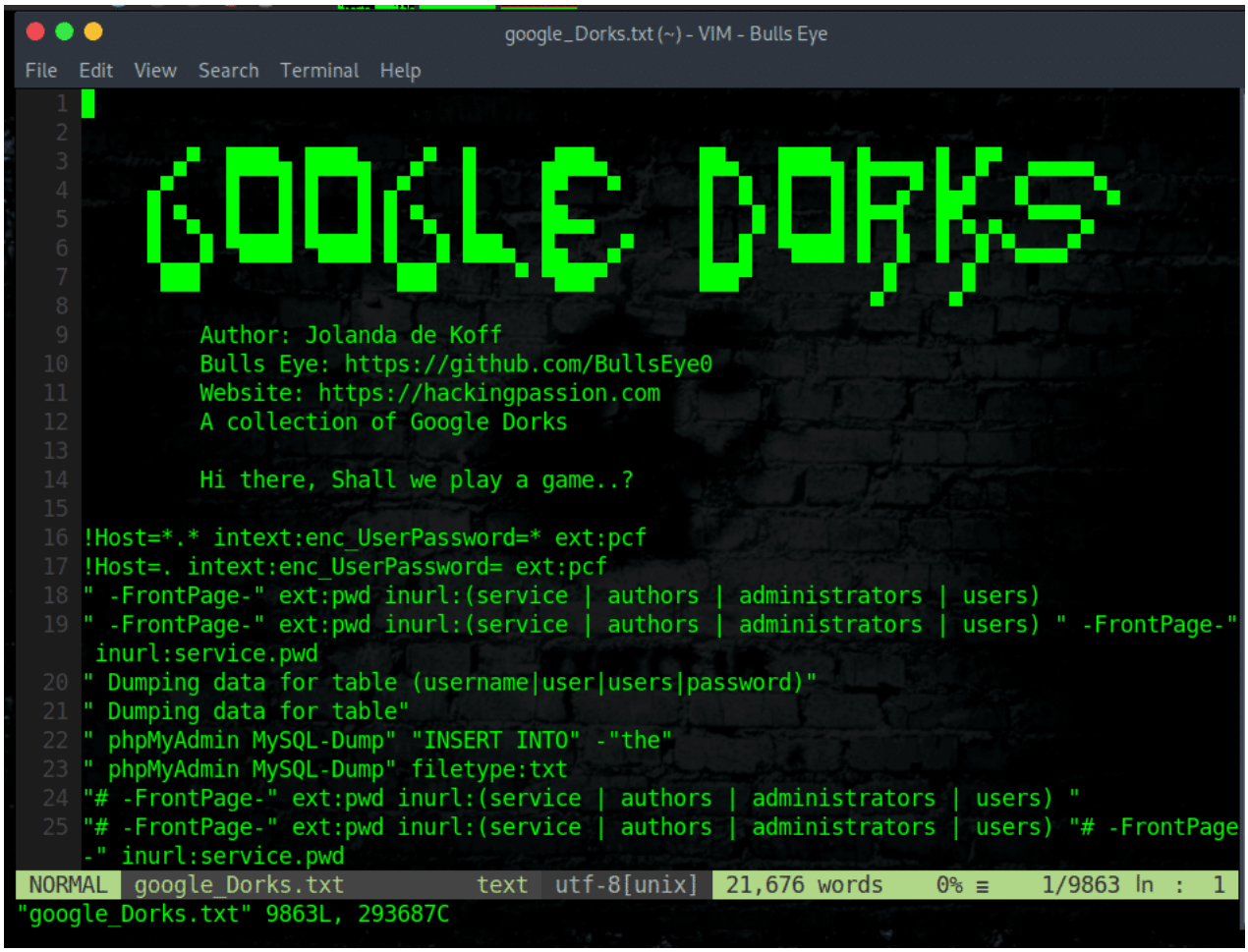

Рисунок 2.6 Визначення області пошуку

TweetDeck є веб-додатком для відстеження та аналізу X, спеціально розробленим для активних користувачів платформи. Основні функції TweetDeck для OSINT включають моніторинг твітів та хештегів у реальному часі, планування публікацій, вивчення трендів, можливості фільтрації, створення та використання списків, а також взаємодію з різними аспектами X (рисунок 2.7). При використанні TweetDeck важливо дотримуватися етичних та правових

стандартів, ураховувати обмеження та ризики, а також забезпечувати захист приватності та прав осіб, які можуть бути задіяні в дослідженні. Окрім переваг, таких як розширені можливості моніторингу та ефективне вивчення трендів, TweetDeck має обмежені аналітичні можливості, залежність від X API, а також питання безпеки та конфіденційності даних користувачів. Перевірка та виконання умов використання X, урахування змін у політиці та правилах, а також відповідальне використання для OSINT-цілей допомагають максимально використовувати можливості TweetDeck у межах закону та етики [44].

Рисунок 2.7 Моніторинг твітів у реальному часі

Hunchly — це інструмент для відкритого джерела інформації (OSINT), спеціально розроблений для збору, відстеження та аналізу великої кількості даних з Інтернету. Функції та характеристики включають автоматичний збір даних, збереження та організацію інформації, відстеження змін, експорт та обмін даними, інтеграцію з іншими інструментами, забезпечення конфіденційності та безпеки (рисунок 2.8). Переваги Hunchly включають ефективність, секурити, гнучкість та контроль, а недоліки — платний сервіс та необхідність навчання. Користувачам слід ураховувати правові обмеження, дотримуватися умов використання, дбати про конфіденційність, уникати недозволених дій та розуміти легальність своїх дій у своїй країні. Hunchly використовується з дотриманням законності та етичних стандартів для збору та аналізу відкритої інформації [45].

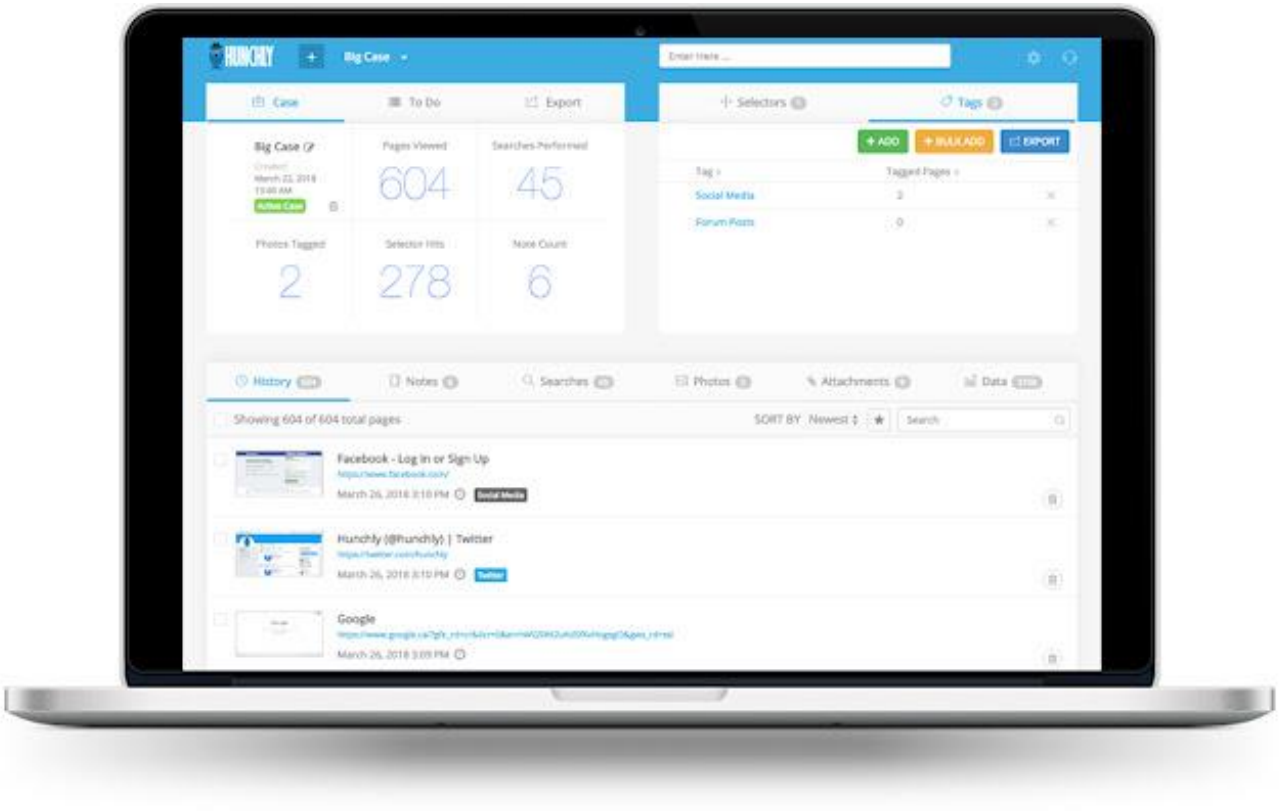

Рисунок 2.8 Відстеження та аналіз автоматичного збору даних

SpiderFoot - інтелектуальний інструмент для збору та аналізу відкритих джерел (OSINT), спрямований на автоматизоване виявлення та аналіз інформації про цільові об'єкти. Функції включають автоматизований збір даних з різних джерел, аналіз зв'язків, розширені пошукові запити, підтримку проксі, візуалізацію результатів, розширюваність модулями (рисунок 2.9). SpiderFoot надає комплексну інформацію, ефективність через автоматизацію та можливість візуалізації результатів. Недоліки включають складність для новачків та обмеженість власних джерел. При використанні інструментів OSINT, важливо дотримуватися законів про конфіденційність, уникати порушення авторського права та враховувати закони про кібербезпеку. Етичне та добросовісне використання завжди рекомендується для уникнення негативних наслідків та порушень закону [48].

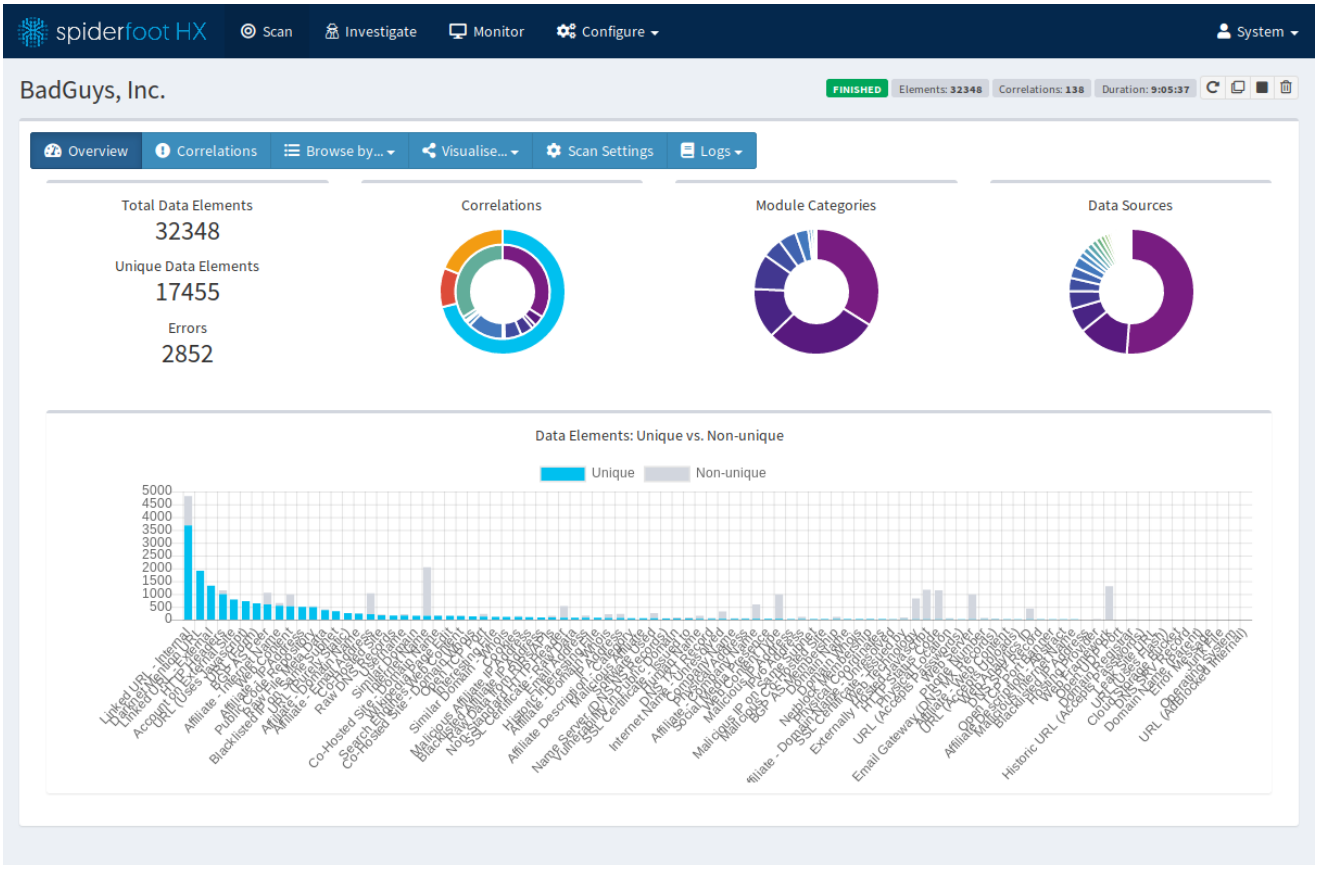

Рисунок 2.9 Візуалізація результатів автоматизованого збору даних

Creepy - це графічний інтерфейс для збору географічної інформації з соціальних мереж, спрямований на аналіз та візуалізацію геолокаційних даних, що користувачі викладають у відкритий доступ (рисунок 2.10). Основні функції включають збір геолокаційних даних з соціальних мереж, візуалізацію результатів, аналіз соціальних мереж, та підтримку різних джерел даних, таких як Twitter та Flickr. Використання Creepy, як і інших інструментів OSINT, повинно відповідати відповідним законам та етичним стандартам, оскільки неправомірне використання може порушити права та приватність осіб. Creepy дозволяє візуалізувати геодані на карті, працює з різними соціальними мережами, має гнучкі налаштування та є відкритим програмним забезпеченням. Недоліки включають обмежену підтримку соціальних мереж та можливість втрати доступу через API. Використання Creepy вимагає уважного відношення до законодавства та етичних норм, оцінку його ефективності та свідоме врахування обмежень і можливих наслідків [47].

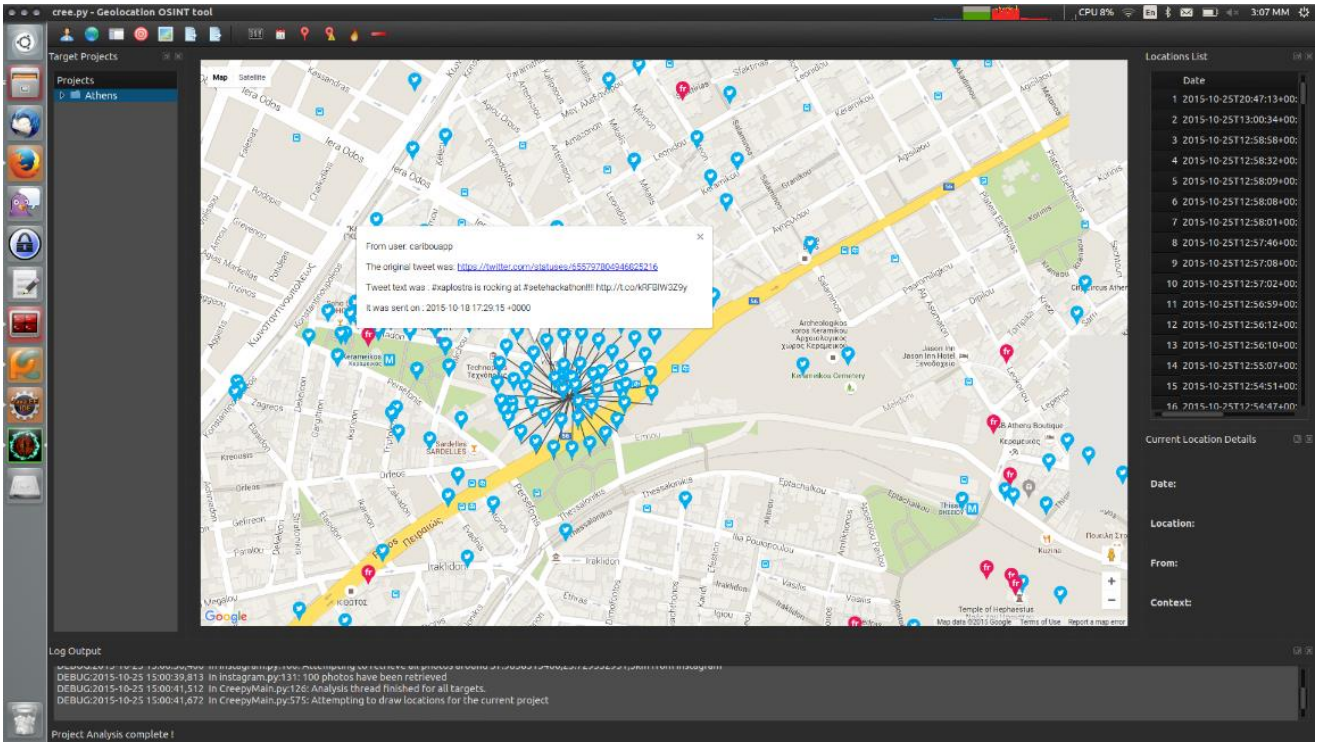

Рисунок 2.10 Скріншот вікна збирання геолокаційних даних

IntelTechniques - набір інструментів для аналізу джерел відкритої інформації (OSINT). Розробник, Майкл Базел (Мозаїч), відомий в галузі кібербезпеки та OSINT. Функції включають пошук інформації в Інтернеті, аналіз електронної пошти, пошук по особі, аналіз зображень та інші. Важливо враховувати, що інструмент може оновлюватися, тому рекомендується перевіряти офіційні джерела для останніх деталей. Переваги IntelTechniques включають широкий функціонал, ефективний пошук, постійне оновлення та активну спільноту користувачів (рисунок 2.11). Недоліки включають обмеження доступу, необхідність навичок, проблеми з правовою стороною та залежність від актуальності даних. Основні правові аспекти стосуються конфіденційності, авторського права, доступу до інформації та етичного використання [30].

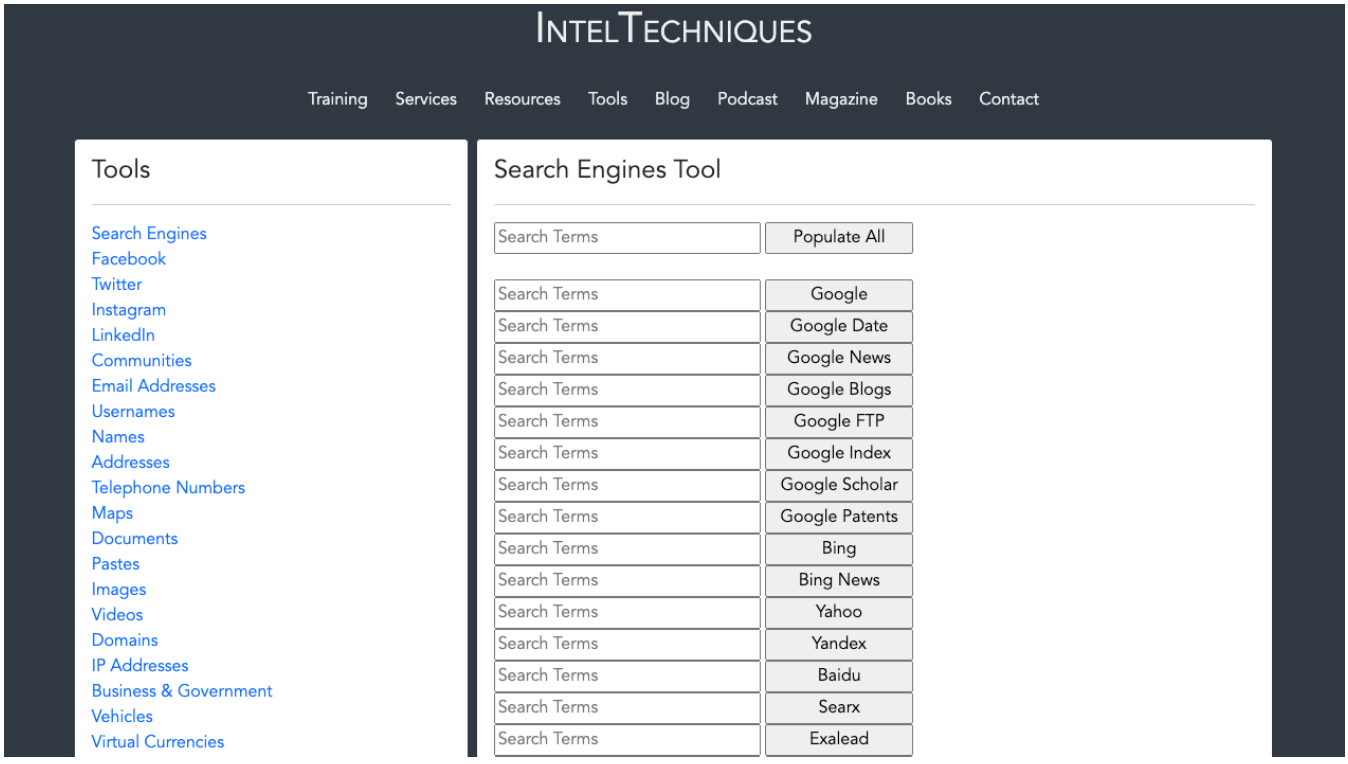

Рисунок 2.11 Інструменти для пошуку інформації та її аналізу

theHarvester - інструмент для аналізу джерел відкритої інформації (OSINT), спроектований для збору електронних адрес та іншої інформації. Його характеристики включають здатність знаходження електронних адрес через різноманітні джерела, такі як веб-сайти та соціальні мережі, підтримку різних джерел, таких як Google, Bing, LinkedIn, PGP keyservers та інші, можливість встановлення параметрів фільтрації для зручного збору конкретної інформації, генерацію звіту зі зібраною інформацією та підтримку використання API для отримання даних (рисунок 2.12).

Переваги theHarvester включають широкий спектр джерел, легкість використання, параметри фільтрації, підтримку API та відкритий вихідний код. Недоліки включають обмеженість джерел, можливі зміни у відкритих джерелах, необхідність у віртуозному використанні, відсутність автоматичної обробки CAPTCHA, можливі проблеми з конфіденційністю та авторським правом, а також необхідність дотримання законодавчих вимог [36].

Фахівці повинні бути уважними до етичних питань, таких як захист приватності осіб та відповідність законам про конфіденційність та авторське право. Важливо також дотримуватися правил використання API та уникати несанкціонованого доступу до систем.

Рисунок 2.12 Можливості функціоналу theHarvester

OSINT Framework є засобом для фахівців з відкритого аналізу джерел відкритої інформації (OSINT), призначеним для зручного доступу до різноманітних інструментів та ресурсів, для збору та аналізу інформації в Інтернеті []. Основні характеристики включають категорії інструментів, такі як соціальні мережі, пошукові системи, аналіз доменів, аналіз електронної пошти, інструменти для компаній, аналіз зображень і відео, а також освітні ресурси (рисунок 2.13) [23].

Переваги OSINT Framework полягають у широкому спектрі ресурсів, їхній оновленості, освітніх матеріалах та універсальності. Недоліки включають неоднозначність ефективності, складність використання, необхідність вивчення кожного інструменту, актуалізацію інформації та правові обмеження [39].

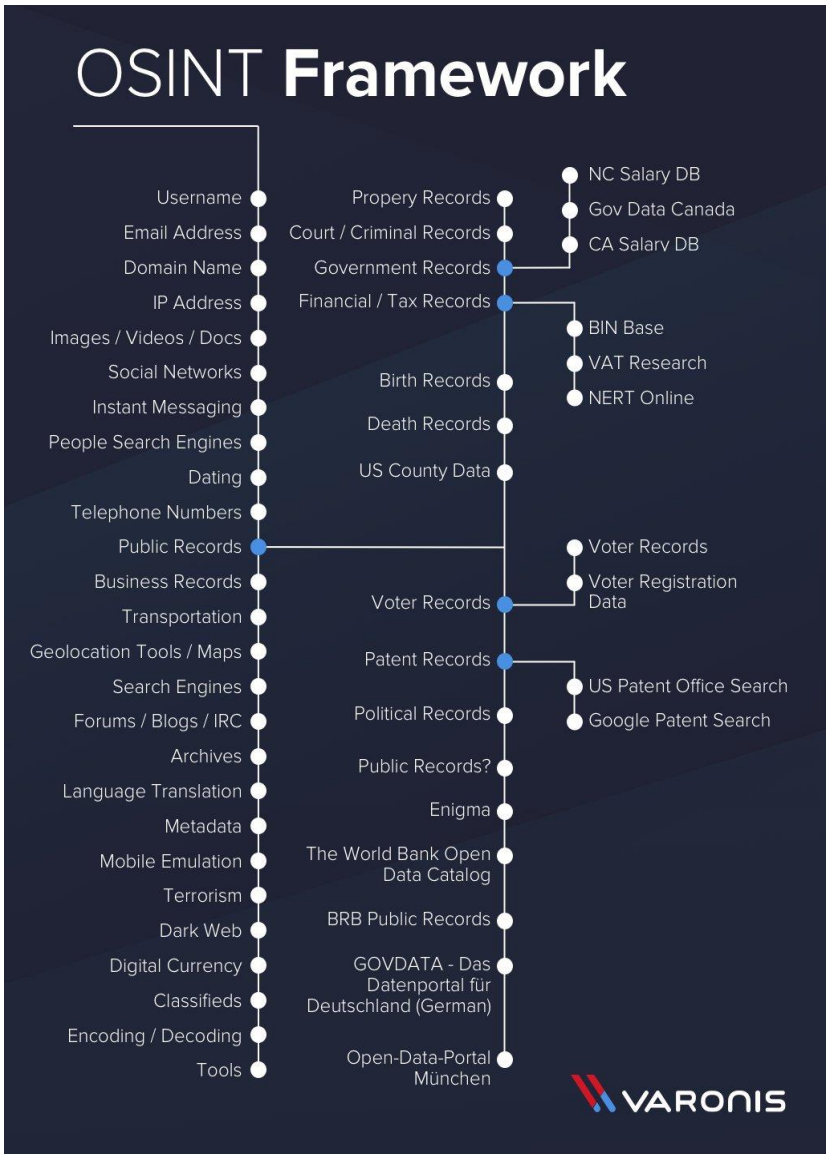

Рисунок 2.13 Категорія інструментів та їх використання

Використання інструментів OSINT для організацій не обов'язково вимагає значних витрат [29]. Навіть якщо конкретний інструмент відомий своєю вартістю, можна знайти аналогічні рішення за більш прийнятні ціни або взагалі скористатися безкоштовними альтернативами. Наприклад, Maltego пропонує умовно безкоштовну версію з базовим функціоналом, а повна версія доступна за плату. Пошук інформації в Maltego здійснюється через API-перетворювачі, і можливість додавання власних перетворювачів робить його ефективним і без додаткових витрат. Також, варто враховувати наявність відкритого програмного забезпечення для OSINT на "github.com", де користувачі можуть безкоштовно користуватися та вдосконалювати його за власними потребами. Пошук таких інструментів можна провести за запитом "github osint" [60].

Висновки до розділу

Результати дослідження методологій роботи в галузі OSINT підкреслюють необхідність систематизації та ефективного використання відкритих джерел інформації. Структурований підхід до збору та аналізу даних дозволяє OSINT-розвідникам досягати точних та комплексних результатів. Використання методологій, що враховують етичні та правові аспекти, підвищує професіоналізм в сфері OSINT та забезпечує відповідність законодавству. Аналіз програмних інструментів у галузі OSINT підтверджує вагомий внесок технологій у вдосконалення розвідувальних процесів. Використання інструментів, таких як аналітичні платформи, веб-скрапінг та аналіз соціальних мереж, розширює можливості отримання інформації. Оптимальний вибір та комбінація програмних засобів дозволяють забезпечити високу швидкість та точність обробки даних, а також підвищити ефективність розвідувальних операцій в цифровому середовищі.

РОЗДІЛ 3.
РОЗРОБКА СИСТЕМИ АНАЛІЗУ ТА КЛАСИФІКАЦІЇ РЕЗУЛЬТАТІВ ПОШУКУ ІНФОРМАЦІЇ

3.1. Засоби реалізації програмного продукту

Для розробки програмного продукту були використані різноманітні технології, такі як пошукова система, нереляційна система управління базами даних, фреймворк для створення графічного інтерфейсу користувача, та інші інструменти.

Мова програмування C++

Для розробки програмного продукту використано мову програмування C++. Ця мова є однією з найпоширеніших та найшвидших мов високого рівня. Стандартна бібліотека мови пропонує різноманітні інструменти для комфортної розробки. Мова підтримує фреймворк Qt для створення графічного інтерфейсу. C++ є компільованою, статично типізованою мовою загального призначення, що об'єднує властивості низькорівневих і високорівневих мов, але вважається високорівневою. Стандартна бібліотека мови включає алгоритми, регулярні вирази, STL-контейнери, підтримку багатопоточності та інші функції. Зазначено, що C++ успадковує синтаксис від мови C, але не включає всі її можливості.

Мова програмування Python

Для ефективної взаємодії з базою даних Neo4j був створений модуль на мові програмування Python. Python – високорівнева, загальнопризначена мова програмування з динамічною типізацією, автоматичним управлінням пам'яттю та використанням відступів для визначення блоків коду. Ця інтерпретована мова може працювати на різних операційних системах і підтримує як процедурне, так і об’єктно-орієнтоване програмування.

Недоліки Python включають меншу швидкість та вище споживання оперативної пам'яті. Розроблення мови почалося наприкінці 1980-х років як удосконалення мови ABC. Python 2.0, випущений у 2000 році, вніс інновації, такі

як система збору сміття. З випуском Python 3.0 у 2008 році мова значно розширила свої можливості, що призвело до несумісності коду Python3 із Python2 і, відповідно, адаптації бібліотек до нового стандарту.

Qt Creator

Для розробки програмного продукту використовувався редактор Qt Creator, спеціально призначений для роботи з фреймворком Qt. Застосування редактора розширюється його можливостями, і він знаходить застосування навіть без фреймворку. Qt Creator надає унікальний інструмент для створення графічного інтерфейсу [52].

Основне призначення Qt Creator полягає в полегшенні розробки програм на різних платформах за допомогою фреймворка Qt. Серед його можливостей є загальні функції середовища розробки та специфічні, такі як налагодження QML-додатків та відображення даних з контейнерів Qt. Редактор включає вбудований дизайнер інтерфейсів, а також використовує систему автоматичного збирання проекту qmake та підтримує git та інші системи контролю версій.

Фреймворк Qt

Для створення кросплатформного графічного інтерфейсу користувача використано фреймворк Qt, який базується на мові програмування C++. Qt є повністю об’єктно орієнтованим і дозволяє розробляти платформо-незалежне програмне забезпечення для Linux, Windows і macOS. Фреймворк включає класи для мережевого програмування, роботи з базами даних та створення графічного інтерфейсу, а також надає готові елементи для побудови інтерфейсу. У даній роботі використані інструменти Qt, зокрема [53]:

- QtNetwork для мережевого програмування та відправлення HTTP запитів;
- QJson для роботи з JSON файлами;
- QGraphicsScene для створення графічних сцен і елементів.

Для збереження даних використано графову систему управління базами даних Neo4j, яка є нереляційною СУБД і однією з популярних графових баз даних. Дані зберігаються у спеціальному форматі для графової інформації, що дозволяє застосовувати оптимізацію при виконанні запитів до даних зі зв’язками.

Використана мова запитів – Cypher. Neo4j також відрізняється оптимізацією обробки графових даних, дозволяючи ефективно обробляти великі графи, не повністю завантажуючи їх в оперативну пам’ять.

Elasticsearch

Для виконання пошуку використана система Elasticsearch, одна з найпопулярніших пошукових систем. Elasticsearch забезпечує швидкий та реального часу доступ до великих обсягів даних. У складі системи використані інструменти, такі як Logstash для збору, перетворення та збереження даних, Kibana - веб-інтерфейс для взаємодії з даними в індексах, та Beats - агент на серверах для відправки різних типів даних в Elasticsearch [54].

Elasticsearch виконує токенізацію, видалення допоміжних символів та інші операції для первинної обробки даних. Після цього відбувається індексація, де дані зберігаються у вигляді документів JSON. Кожен документ містить набір ключів та відповідних значень. Elasticsearch визначає унікальні слова в документах та ідентифікує всі документи, що містять ці слова. Такий підхід оптимізує процес пошуку великих обсягів даних.*Фреймворк*

Graphaware

Для синхронізації пошукової системи та бази даних Neo4j використовувався фреймворк Graphaware. Цей фреймворк з відкритим вихідним кодом забезпечує тісну взаємодію між Neo4j та Elasticsearch, розширюючи функціональність бази даних та забезпечуючи синхронізацію даних у реальному часі з мінімальною затримкою.

Бібліотека Beutifulsoup

Beautiful Soup, написаний на мові програмування Python, є парсером для синтаксичного розбору файлів HTML/XML. Він здатен перетворити неправильну розмітку в дерево синтаксичного розбору та надає прості методи навігації, пошуку та модифікації цього дерева. Застосовується для розбору HTML сторінок та вилучення інформації з них.

3.2. Опис програмної реалізації

Архітектура програмного коду

Система інформаційного пошуку на основі семантичної мережі використовує Beutifulsoup для отримання даних з веб-сторінок, зберігає їх у базі даних Neo4j у вигляді онтологічної моделі і використовує Elasticsearch для виконання пошуку. Синхронізація даних з бази даних в пошукову систему забезпечується, а пошук враховує створені зв'язки семантичної мережі. Система складається з шести модулів: модуль збору даних, модуль відправки даних до бази даних, модуль інтерфейсу, модуль синхронізації бази даних, пошукова система та графова база даних [57].

Модуль збору даних

Модуль збору даних завантажує HTML код веб-сторінки, вилучає текстову інформацію для подальшого пошуку в системі за допомогою бібліотеки Beutifulsoup. Після збереження HTML сторінки формується синтаксичне дерево, з якого вилучається інформація з типом "text". Також проводиться завантаження всіх посилань на інші сторінки, визначених у синтаксичному дереві, забезпечуючи збір інформації з веб-сторінки та всіх пов'язаних посилань.

Модуль відправки даних до БД

Зібрана інформація подається у модуль бібліотеки Neo4j для відправки даних в базу даних за допомогою мови запитів Cypher. Приклади Cypher запитів включають створення елементу "Website" з визначеними полями, створення зв'язку з назвою "MENTIONED" між елементами та видалення всіх елементів та зв'язків за певних умов.

Модель інтерфейсу

Модуль інтерфейсу – це користувацький інтерфейс, реалізований на мові програмування C++ з використанням фреймворку Qt. Інтерфейс має поля для введення пошукового запиту, кнопку для введення веб-сторінки в базу даних та поле для відображення результатів пошуку.

При натисканні на кнопку "Specify URL" користувач може ввести адресу веб-сторінки для збереження в базі даних. Після завершення пошуку користувач може вибрати знайдену сторінку, а програма автоматично відкриє її в веб-браузері.

Модуль синхронізації бази даних

Фреймворк Graphaware виконує синхронізацію отриманої базою даних інформації з пошуковою системою Elasticsearch.

Пошукова система

Система Elasticsearch індексує та обробляє дані, після чого модуль інтерфейсу отримує результати пошуку за конкретним запитом. Синхронізація даних з базою відбувається через відповідний модуль. Запити POST використовуються для отримання результатів пошуку з вказаною структурою, де "UserRequest" – пошуковий запит.

База даних

Графова база даних Neo4j використовується для збереження інформації, що дозволяє створювати зв'язки між даними та впливати на результат пошуку. В базі даних зберігаються веб-посилання під ключем "title" та інформація веб-сторінки під ключем "content". Зв'язок "MENTIONED" формується між основною веб-сторінкою та тією, яка знайдена серед посилань всередині основної.

Алгоритм роботи програми

Програма дозволяє користувачеві завантажувати дані з відкритих джерел і виконувати інформаційний пошук, використовуючи збережені дані. Користувач вказує веб-посилання, яке передається модулю збору даних через інтерфейс. Запит містить веб-адресу веб-сайту для зчитування та збереження в базу даних у форматі "string" [58].

Модуль збору інформації завантажує вказану веб-сторінку у форматі HTML та створює синтаксичне дерево. З дерева вилучається інформація з типом "text" та веб-посилання з типом "a-href". Отримані посилання використовуються для збору інформації з типом зв'язку "MENTIONED" в базі даних [61].

Користувач може також завантажити дані у форматі текстових файлів, які модуль збору інформації зчитує та зберігає в оперативну пам'ять, включаючи інформацію про створені зв'язки. Зібрана інформація передається модулю відправки даних до бази даних, який використовує мову Cypher для створення елементів Neo4j та додавання зв'язків [62].

Після завершення транзакцій дані зберігаються у базі даних у формі семантичної мережі. Модуль синхронізації отримує сигнал від бази даних про нову інформацію і передає її в пошукову систему .

Модуль пошукової системи індексує отриману інформацію та готує її для пошуку. Модуль інтерфейсу передає пошуковий запит в пошукову систему та отримує релевантні результати. Крім того, користувач може відкривати веб-посилання з результатів пошуку в веб-браузері.

3.3. Розробка програмного продукту

На початковому етапі передбачено розроблення графічного інтерфейсу за допомогою Qt Creator, багатоплатформового інтегрованого середовища програмування мовою C++, спеціалізованого для використання Qt. Існує можливість інтеграції Qt з Code Blocks. Це ефективний інструмент для створення графічних інтерфейсів, що спрощує адаптацію програми на різних платформах, дозволяючи зосередитися на задачах та алгоритмах розв'язання завдань, а не на проблемах сумісності з операційною системою (рисунок 3.1).

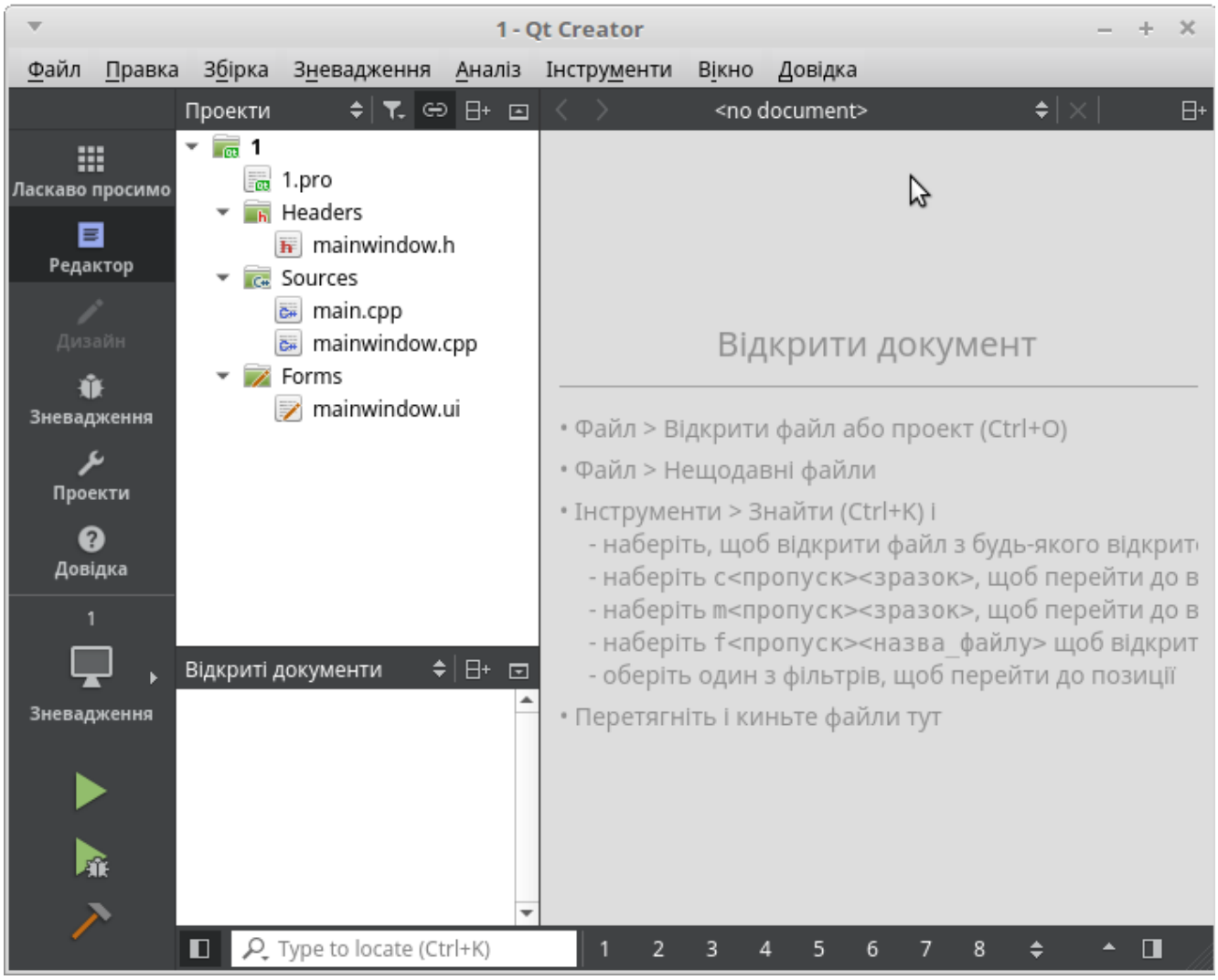

Рисунок 3.1 Розробка графічного інтерфейсу

Наочне редагування форми здійснювалось на вкладенні Design (Дизайн), що надає доступ до великої кількості елементів керування (рисунок 3.2).

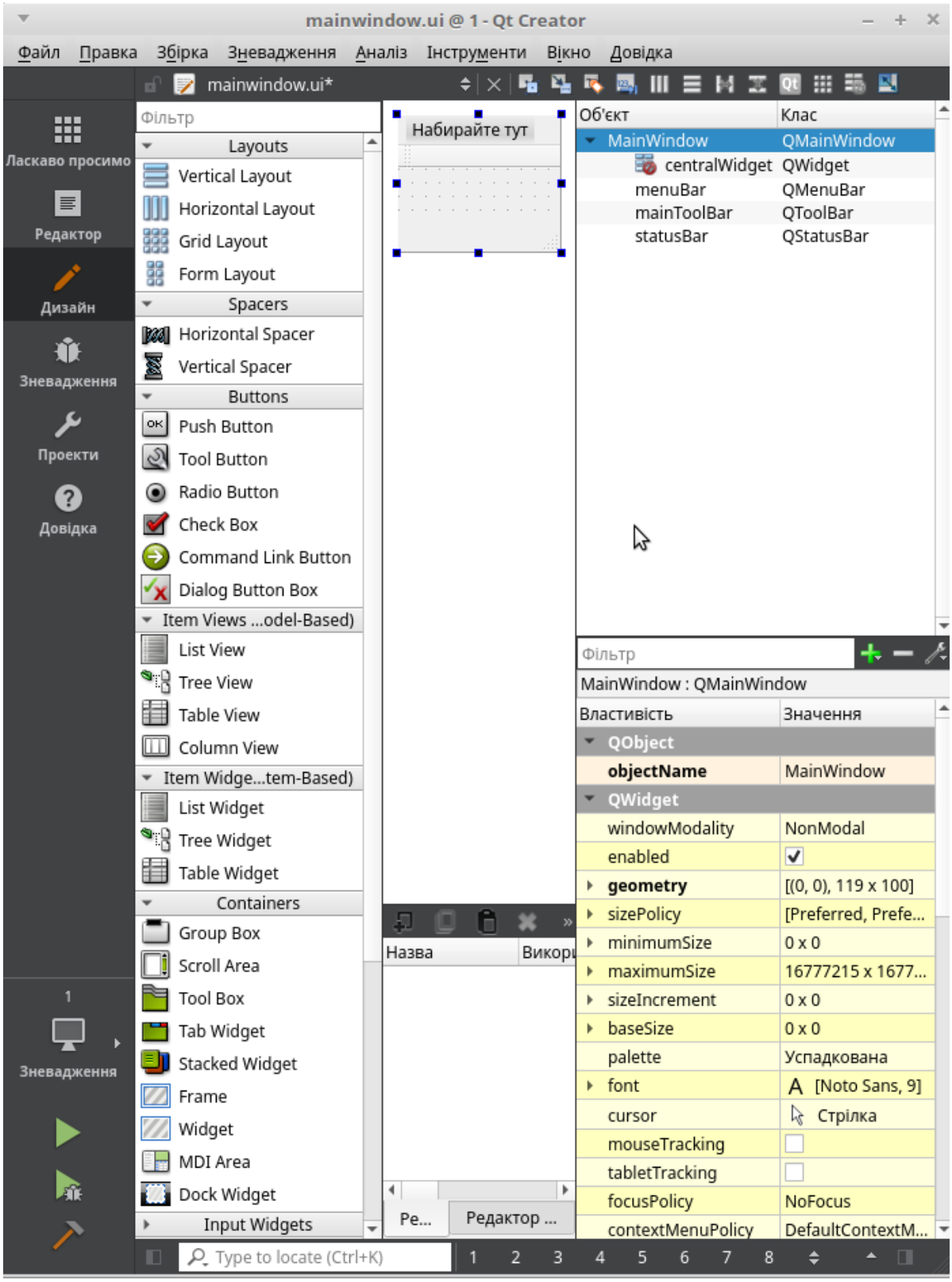

Рисунок 3.2 Наочне редагування форми

Детальний опис роботи поданий у підрозділах (3.1, 3.2) та (Додаток Г).

Під час другого етапу розробки графової бази даних використовувалася система управління базами даних NEO4J, що є графовою СУБД з відкритим

вихідним кодом, реалізованою на Java. Ця система є провідною в графовому напрямку та має аналоги у вигляді Oracle NoSQL Database, HypherGraphDB, GraphBase, InfiniteGraph та AllegroGraph (рисунок 3.3).

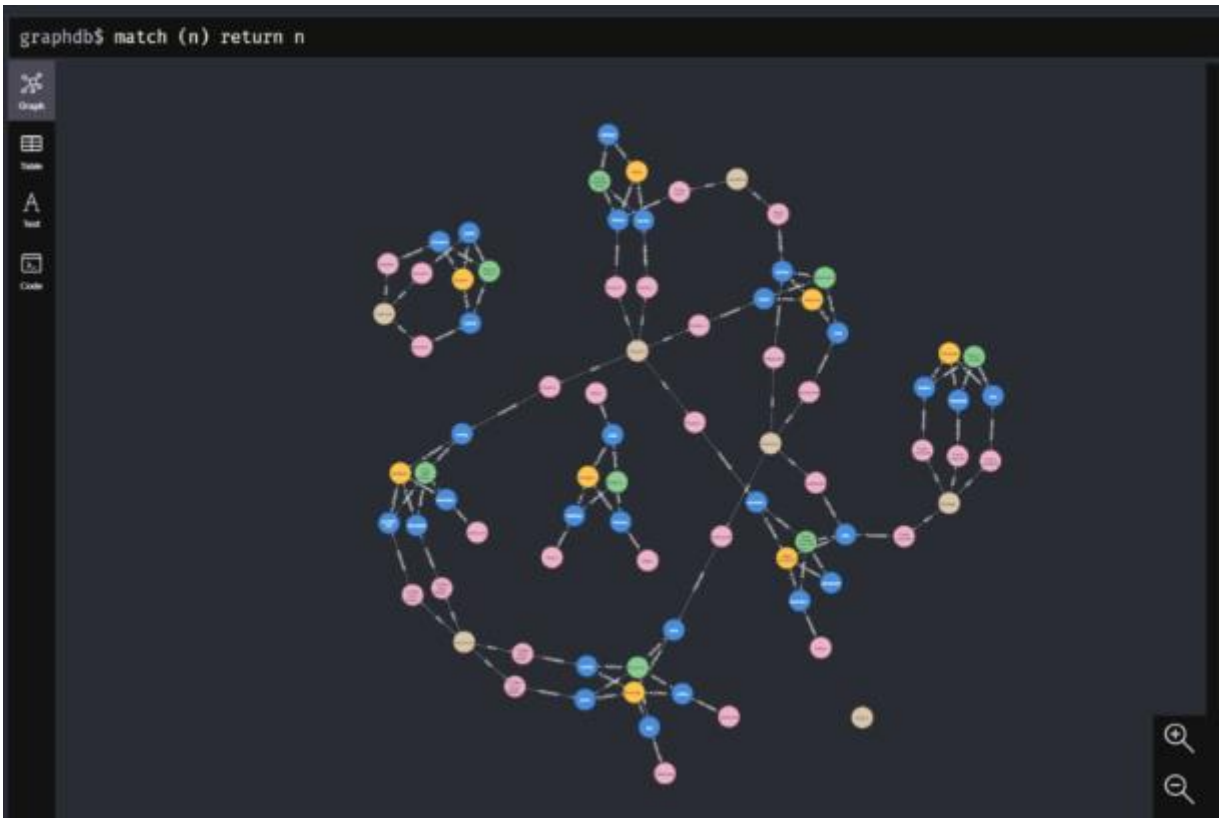

Рисунок 3.3 Розробка графової бази даних

Детальний опис наведено у підрозділах (3.1, 3.2) та (Додаток Б).

На третьому етапі використано систему Elasticsearch, одну з провідних пошукових систем, яка забезпечує швидкий та реального часу доступ до великих обсягів даних (рисунок 3.4).

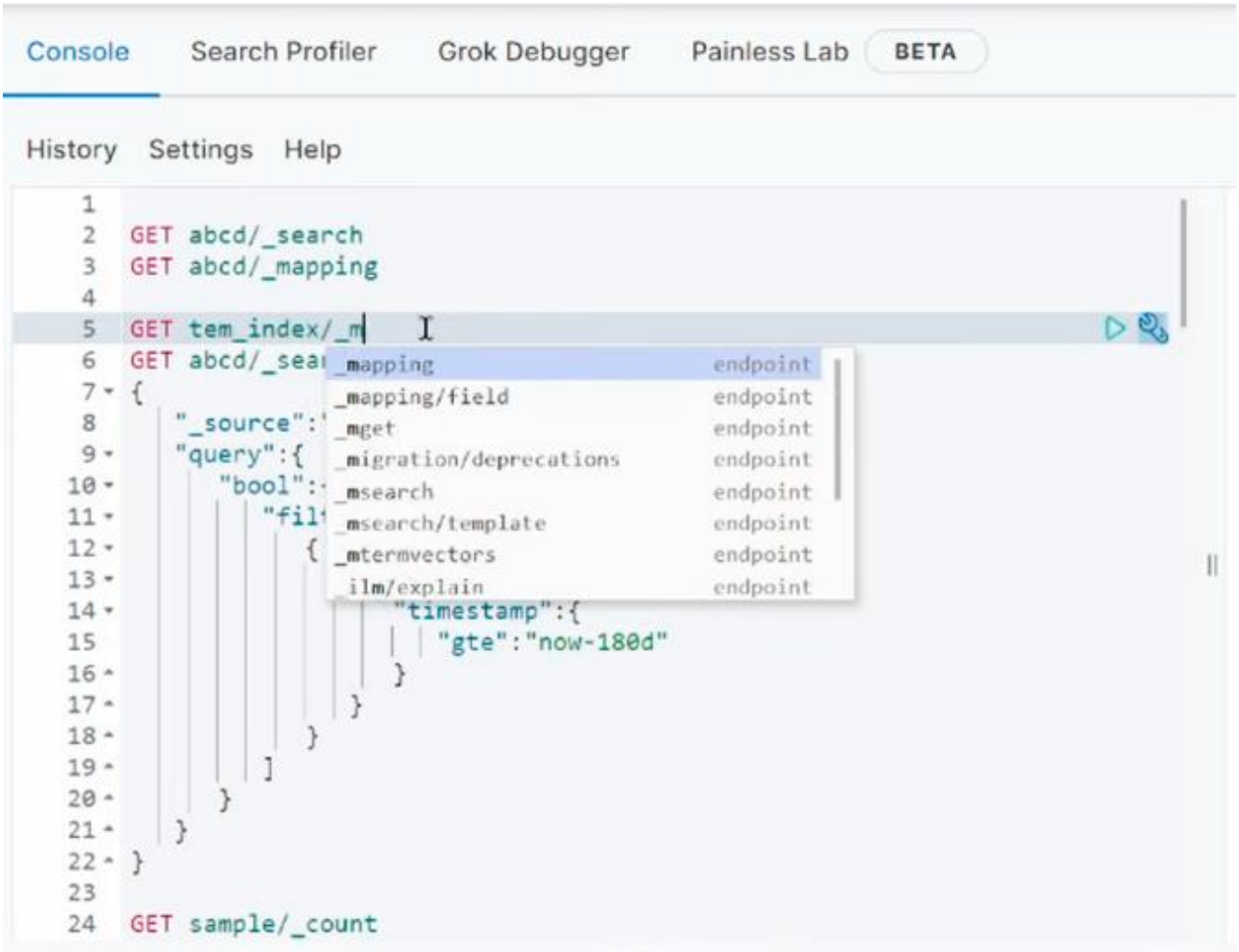

Рисунок 3.4 Початок роботи з системою Elasticsearch

Детальний розгляд наведений у підрозділах (3.1, 3.2) та (Додаток В).

На четвертому етапі впроваджено фреймворк GraphAware для синхронізації бази даних NEO4J з системою Elasticsearch, детальніше дивитися підрозділи (3.1, 3.2) та (Додаток Г).

На п'ятому етапі представлений результат проведеної роботи. Користувацький інтерфейс дозволяє вибрати веб-сторінку для збереження за допомогою функції "Specify URL".

Натисканням кнопки відкривається вікно для введення веб-адреси користувачем (рисунок 3.5).

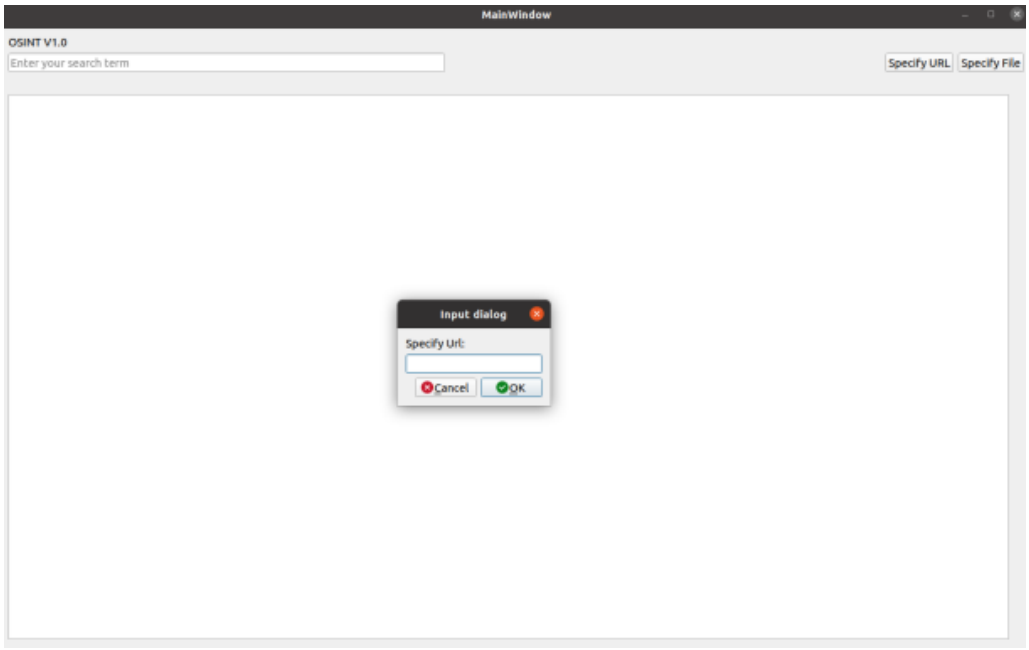

Рисунок 3.5 Вікно вводу веб-сторінки

Користувач може скасувати операцію, натискавши "Cancel". Після введення веб-адреси та натискання "Ok" з'являється вікно попередження про початок операції збереження інформації, під час якої обмежується можливість виконання пошуку або введення нових адрес (рисунок 3.6).

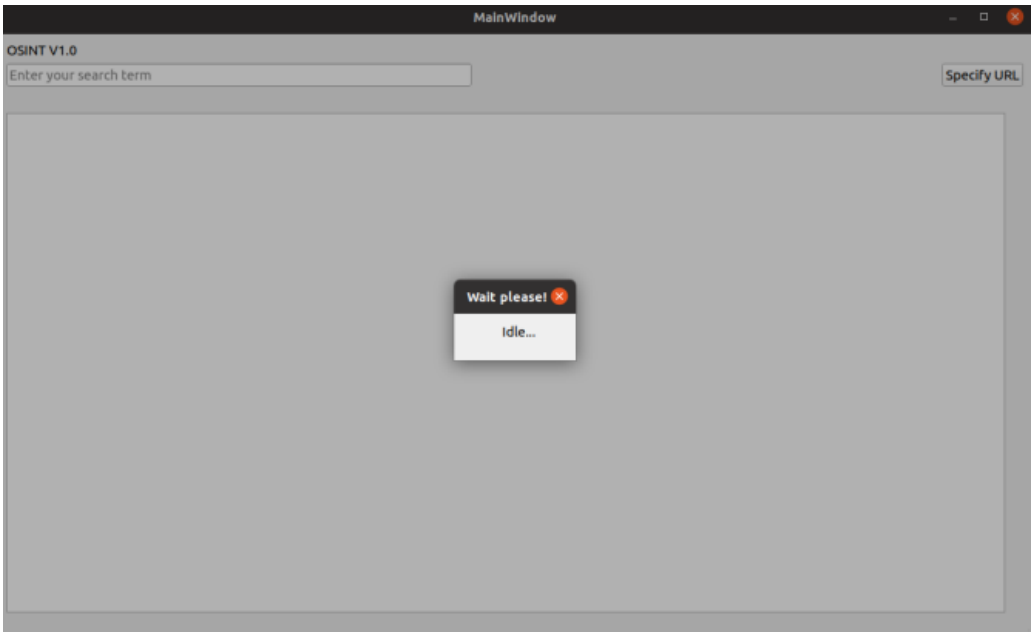

Рисунок 3.6 Вікно очікування

Натисканням кнопки "Specify file" відкривається вікно для збереження інформації з текстових файлів, де користувач може завантажувати дані до бази даних та встановлювати зв’язки між файлами (рисунок 3.7).

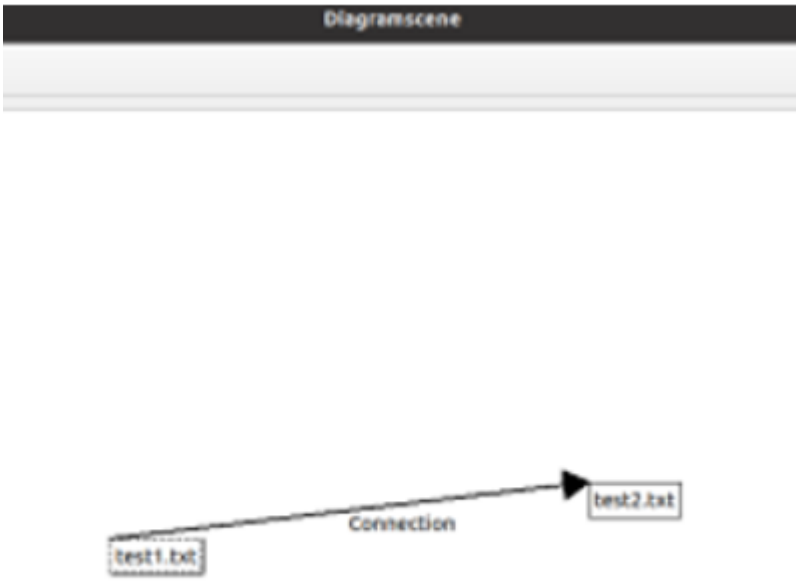

Рисунок 3.7 Вікно збереження текстових файлів

У лівій частині вікна розташовані кнопки управління елементами, де користувач може створювати, видаляти, переміщувати та змінювати масштаб елементів. Кнопка "Add text file" дозволяє користувачу створювати новий елемент шляхом натискання на вільну область у головному вікні (рисунок 3.8).

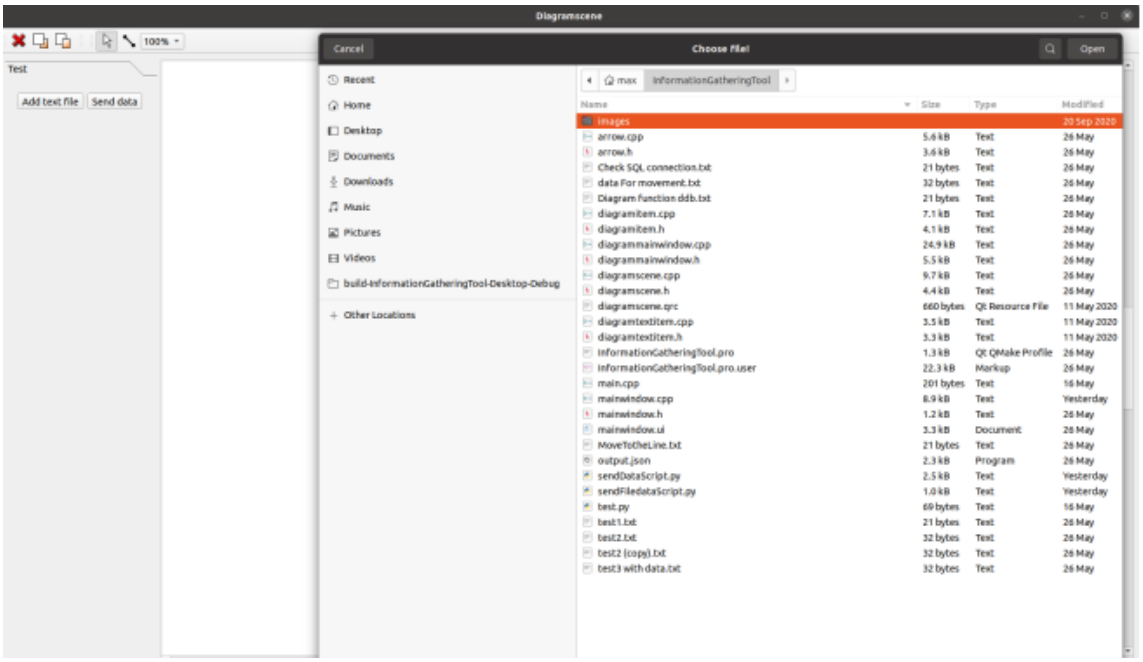

Рисунок 3.8 Вікно вибору текстового файлу

Під час вибору файлу користувач визначає необхідні файли; текстова інформація автоматично зчитується та зберігається в динамічній пам'яті. При невідповідності формату файлу текстовому, виводиться помилка формату.

У верхньому меню доступна опція видалення, яка дозволяє користувачеві видаляти створені елементи .

При наявності великої кількості елементів користувач може їх накладати один на одного. Управління відображенням цих елементів здійснюється за допомогою кнопок "Send to back" та "Send to front" (рисунок 3.9).

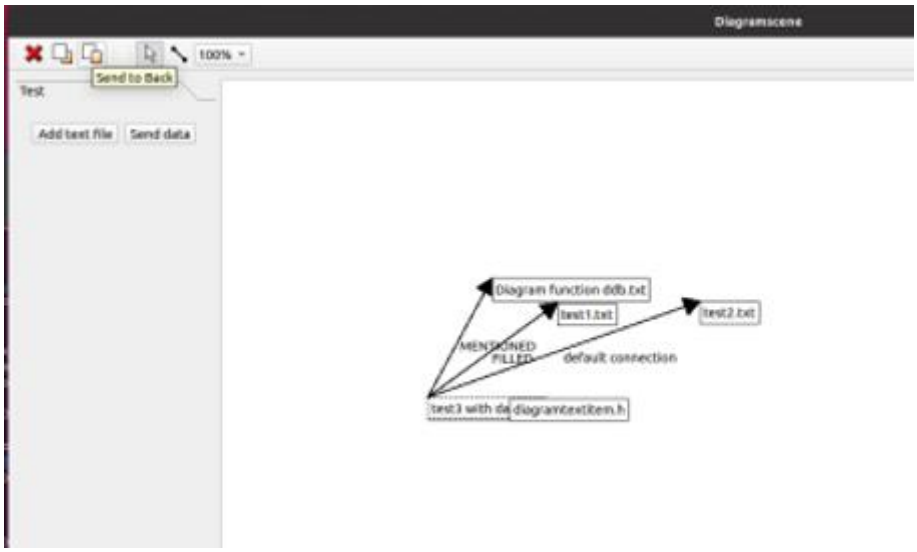

Рисунок 3.9 Керування відображенням елементів

Кнопкою створення зв'язків користувач обирає два елементи, між якими створюються зв'язки. Далі відкривається вікно для введення назви зв'язку (рисунок 3.10).

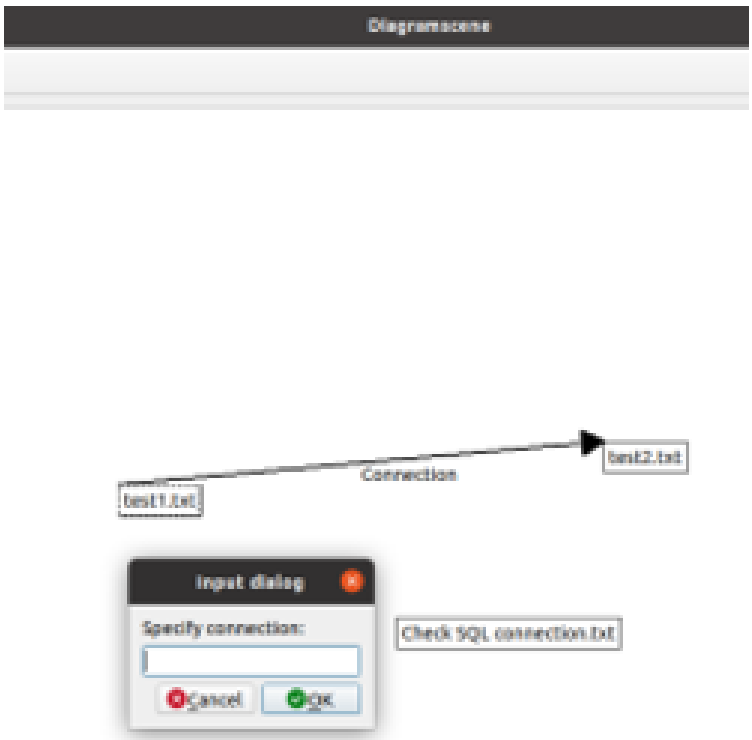

Рисунок 3.10 Вікно створення зв’язків

Завершивши операцію збору даних, користувач може використовувати систему пошуку, введенням запиту у поле "Enter your search term"

Користувач може необмежено здійснювати операції збереження даних перед початком використання системи пошуку. Після натискання клавіші "Enter" виконується запит, і виводяться результати пошуку (рисунок 3.11).

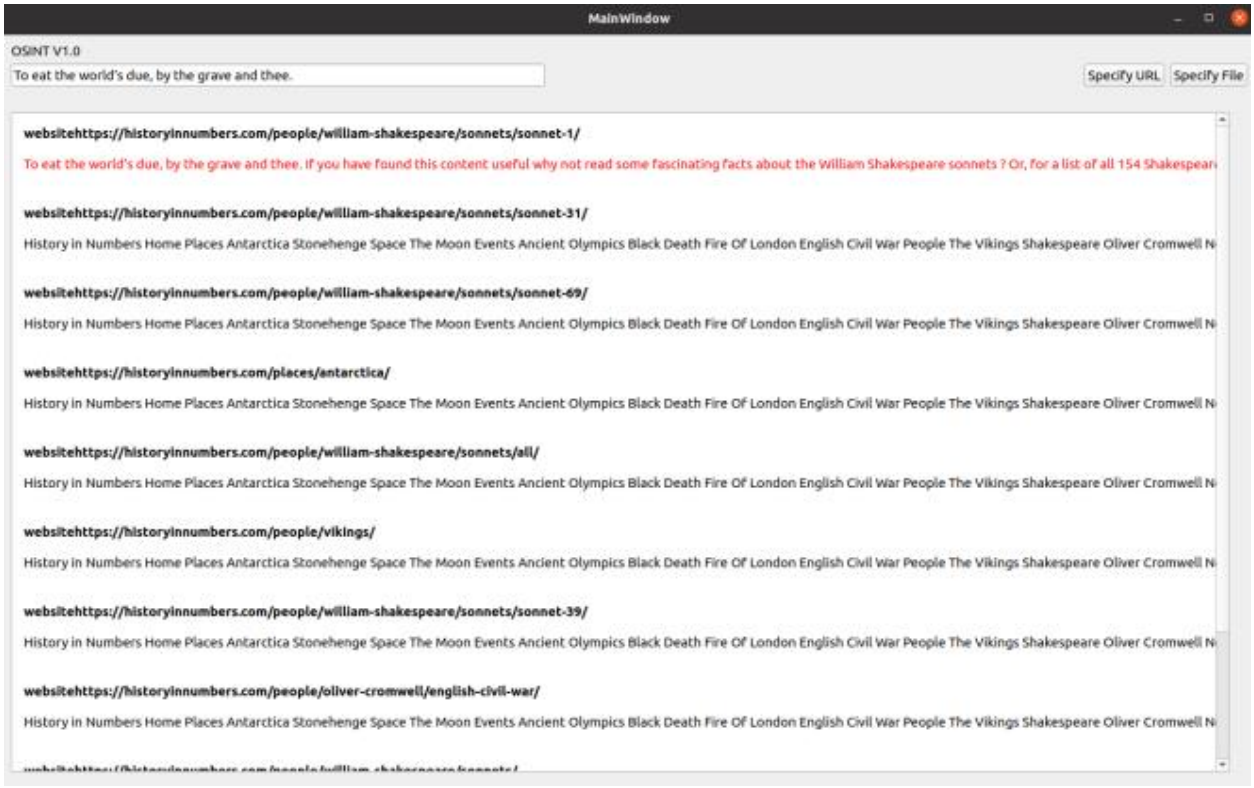

Рисунок 3.11 Результат за запитом “To eat the world, by the grave and thee.”

У результатах запиту відображаються веб-адреса та текст веб-посилання. Точне співпадіння виділяється червоним коліром. Результати пошуку сортуються за значенням релевантності у порядку спадання; результати з релевантністю менше 0.01 вважаються не релевантними й не виводяться. Значення релевантності вказується при наведенні на результат. Користувач може відкрити веб-посилання, натиснувши на елемент результату пошуку у браузері (рисунок 3.12).

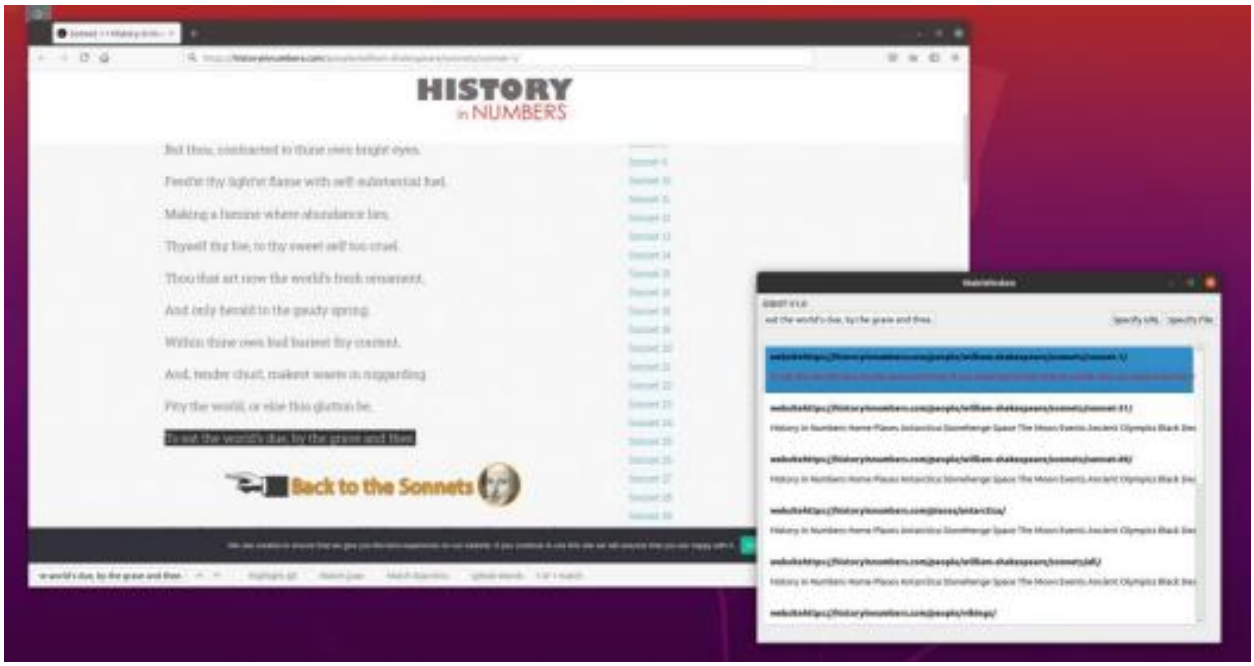

Рисунок 3.12 Відкрите посилання за результатом

Користувач може виконувати необмежену кількість запитів, а інформація, збережена в базі даних, застосовується для подальших пошукових операцій.

Висновки до розділу

У цьому розділі розглянуто використані мову програмування та технології для розробки програмного продукту. Для графічного інтерфейсу обрано фреймворк Qt, з використанням Qt Creator як середовища розробки. В якості бази даних обрано нереляційну базу Neo4j, а для інформаційного пошуку використана система Elasticsearch. Розділ також містить опис архітектури, компонентну схему, діаграму прецедентів і структуру бази даних. Програмний продукт складається з шести модулів: збір даних, відправка даних до БД, інтерфейс, синхронізація БД, графова база даних та пошукова система. Розроблено і впроваджено нереляційну базу даних, яка зберігає інформацію у вигляді семантичної мережі. Алгоритм роботи програми описано, включаючи зчитування та збереження даних з відкритих джерел у базу даних та їх використання для формування релевантних результатів пошуку з використанням семантичних зв’язків.

ВИСНОВКИ

В результаті виконання кваліфікаційної роботи було досягнуто мету- досліджено різноманітні засоби пошуку та аналізу інформаційних ресурсів з відкритих онлайн-джерел, вивчено можливості їхнього оптимального застосування, а також визначення перспектив розвитку OSINT-технологій у майбутньому. Виконано всі завдання дослідження, а саме:

1. Вивчено основні поняття та принципи OSINT-технології.
2. Проаналізовано основні сфери застосування OSINT- технології.
3. Досліджено різноманіття технічних, програмних та інформаційних ресурсів, що використовуються у сфері OSINT.
4. Проаналізовано ефективність багатьох методів та інструментів для збору відкритої інформації.
5. Спроектовано та розроблено систему, що дозволяє ефективно аналізувати та класифікувати результати пошуку відповідно до визначених критеріїв.

В ході та за підсумками виконання цієї кваліфікаційної роботи мною були зроблені такі висновки: результати дослідження підтверджують, що OSINT є ключовою складовою розвідки, забезпечуючи безпеку та ефективні розвідувальні заходи через використання публічної інформації; основні методи та інструменти OSINT призначені для отримання об'єктивних та релевантних даних.

Вивчення ключових елементів з використанням публічної інформації є стратегічно важливою частиною розвідувального процесу, що забезпечує глибоке розуміння ситуації та прийняття обґрунтованих рішень. Важливість дотримання етичних та правових норм під час застосування методів OSINT підкреслюється для забезпечення легітимності діяльності.

Розширення сфери застосування OSINT-розвідки ґрунтується на її універсальності в умовах аналізу кібербезпеки та вирішення геополітичних конфліктів, забезпечуючи високий рівень інформаційної обізнаності. Результати

дослідження методологій роботи в галузі OSINT підкреслюють необхідність систематизації та ефективного використання відкритих джерел інформації.

Структурований підхід до збору та аналізу даних дозволяє OSINT-розвідникам досягати точних та комплексних результатів. Використання методологій, що враховують етичні та правові аспекти, підвищує професіоналізм в сфері OSINT та забезпечує відповідність законодавству. Аналіз програмних інструментів у галузі OSINT підтверджує вагомий внесок технологій у вдосконалення розвідувальних процесів. Використання інструментів, таких як аналітичні платформи, веб-скрапінг та аналіз соціальних мереж, розширює можливості отримання інформації.

Оптимальний вибір та комбінація програмних засобів дозволяють забезпечити високу швидкість та точність обробки даних, а також підвищити ефективність розвідувальних операцій в цифровому середовищі.

Для створення системи інформаційного пошуку використовувались наступні інструменти: об’єктноорієнтована мова програмування C++, середовище розробки Qt Creator, СУБД Neo4j, пошукова система Elasticsearch, фреймворк Graphaware, високорівнева мова програмування Python та парсер Beautifulsoup. Розроблено архітектуру програмного продукту, який складається з 6 модулів: модуль збору даних, модуль відправки даних до БД, модуль інтерфейсу, модуль синхронізації БД, графова база даних, пошукова система. Впроваджено нереляційну базу даних, що зберігає інформацію у вигляді семантичної мережі. Порівнюючи із системами Hakia та Kosmix, слід відзначити, що створена система, хоча не має такого обширного функціоналу, пропонує зрозумілий інтерфейс та відсутність реклами.

Розроблений продукт може бути корисним інструментом для інформаційної розвідки, допомагаючи знаходити та зберігати надійні джерела інформації в інтернеті. Використання його як пошукової системи, при умові наявності потрібних даних у базі, може сприяти ефективному інформаційному пошуку. Таким чином, розроблений продукт є актуальним та може бути впроваджений на практиці.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1 Рисюк А. Дослідження технології добування корисної інформації з відкритих онлайн-джерел. Збірник матеріалів наукової конференції здобувачів вищої освіти фізико-математичного факультету Кам'янець-Подільського національного університету імені Івана Огієнка. 1.11.2023 року. Кам'янець-Подільський: Кам'янець-Подільський національний університет імені Івана Огієнка, 2023. С. 40. URL: http://elar.kpnu.edu.ua/xmlui/handle/123456789/7648

2 Andrews RJ. Info We Trust: How to Inspire the World with Data. Wiley. 2019. 272p.

3 A Beginner's Guide to OSINT Investigation with Maltego. URL: https://wondersmithrae.medium.com/a-beginners-guide-to-osint-investigation-with-maltego-6b195f7245cc

4 A Guide To Open Source Intelligence. URL: https://itsec.group/blog-post-osint-guide-part-1.html

5 Academy of cyber technologies of Ukraine. OSINT. Сучасна кіберрозвідка. URL: https://www.youtube.com/watch?v=Qjal2T3lOSU

6 BSides. URL: https://www.securitybsides.com/w/page/12194156/FrontPage

7 Bellingcat. URL: https://www.bellingcat.com

8 Bazzell M., Carroll J. The Complete Privacy & Security Desk Reference: Volume I: Digital. CreateSpace Independent Publishing Platform. 2016. 492 p.

9 Baker R. L. Deep Dive: Exploring the Real-world Value of Open Source Intelligence 1st Edition. Wiley, 2023. 544 p.

10 Brügger N., Schroeder R. Web as History: Using Web Archives to Understand the Past and the Present. UCL Press., Illustrated edition. 2017. 296 p.

11 Bazzell M. Open Source intelligence techniques: Resources for searching and analyzing online information Sixth Edition. Independently published, 2023. 550 p.

12 Bazzell M. Open Source Intelligence Techniques Resources for Searching and Analyzing Online Information. Independently published, 2021. 666 p.

13 Bellingcat’s Digital Toolkit. URL:https://archive.comsuregroup.com/wp-content/uploads/2018/06/Bellingcats-Digital-Toolkit.pdf

14 Blum I., Williams H. J. Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise. Santa Monica: RAND Corporation, 2018. 62 p. URL: https://www.rand.org/pubs/research_reports/RR1964.html

15 Bielska A. Open source intelligence tools and resources handbook. 2020. 510 p. URL: https://i-intelligence.eu/uploads/public-documents/OSINT_Handbook_2020.pdf

16 Everything about Open source intelligence and osint investigations. URL: https://www.maltego.com/blog/what-is-open-source-intelligence-and-how-to-conduct-osint-investigations/

17 Cybernews Blog. URL: https://cybernews.com/blog/

18 COOK S. A Guide to Open-Source Intelligence (OSINT). URL: https://greydynamics.com/a-guide-to-open-source-intelligence-osint/

19 CQR OSINT. URL: https://cqr.company/pentesting-process/osint/

20 Додонов А.Г. Распознавание информационных операций/ А.Г. Додонов, Д.В. Ландэ, В.В. Цыганок, О.В. Андрейчук, С.В. Каденко, А.Н. Грайворонская. Киев: ООО «Инжиниринг», 2017. 282 с.

21 DEF CON. URL: https://www.defcon.org/

22 David Bombal. OSINT social media. URL: https://www.youtube.com/watch?v=F6l2Bmh7Dq4

23 David Bombal Clips. Where to start in OSINT? URL: https://www.youtube.com/watch?v=ALy5bUMUo7Q

24 Goyal S. Sublist3r – Fastest Subdomain Enumeration Tool. URL: https://secnhack.in/sublist3r-fastest-subdomain-enumeration-tool/

25 Maltego - Cyber Weapons Lab - Research like an OSINT Analyst. URL: https://www.youtube.com/watch?v=46st98FUf8s

26 McFarlane D. A Beginners Guide to OSINT. URL: https://www.csnp.org/post/a-beginners-guide-to-osint

27 Hackathon A. OSINT VM. URL: https://www.tracelabs.org/blog/osint-vm-august-hackathon

28 Hadnagy C. Social Engineering. John Wiley &Sons. 2010. 410 p.

29 Mitnick K. Ghost in the Wires: My Adventures as the World's Most Wanted Hacker. Back Bay Books. 2012. 448 p.

30 Hassan N. A., Hijazi R. Open Source Intelligence Methods and Tools: A Practical Guide to Online Intelligence 1st ed. Edition. Apress, 2018. 377 p.

31 IntelTechniques. URL: https://inteltechniques.com/

32 Nicole Beckwith - Mind Hacks – Psychological profiling, and mental health in OSINT investigations. URL: https://www.youtube.com/watch?v=104WpJm_eGk

33 OSINT Curious. URL: https://docs.github.com/en/pages

34 OSINT — Beginner's Guide (Part 1). URL: https://medium.com/@Aardwarewolf/what-is-osint-part-1-91aaa3890643

35 Open Source INTelligence (OSINT) Reference Sheet. URL:https://tpia.com/resources/Pictures/2019%20CPE%20files/OSINT%20Resources.pdf

36 Python theHarvester – How to use it? URL: https://www.geeksforgeeks.org/python-theharvester-how-to-use-it/

37 Open Source Intelligence Techniques (OSINT) for Fraud Prevention. URL: https://seon.io/resources/guides/open-source-intelligence-techniques-osint-for-fraud-prevention/

38 OSINT COMBINE. URL: https://www.osintcombine.com/training

39 OSINT Framework. URL: https://osintframework.com

40 OSINT-розвідка: як дії цивільних людей можуть допомагати ворогу? URL: https://www.youtube.com/watch?v=Gf3wgyykSjM&t=307s

41 OSINT-FR. OSINT Origins. URL: https://www.youtube.com/watch?v=XrTFzZ77eEI

42 OSINT TEAM. The Atypical OSINT. URL: https://osintteam.blog/the-atypical-osint-guide-2023-276a8d00959

43 OSINT Guide – Open Source Intelligence. URL: https://www.osintguide.com

44 Picolet J. Operator Handbook: Red Team + OSINT + Blue Team Reference. Independently published, 2020. 312 p.

45 Top OSINT. URL: https://www.maltego.com/blog/top-osint-infosec-resources-for-you-and-your-team/

46 SANS Cyber Security Webinars. URL: https://www.sans.org/webcasts/

47 SANS Cyber Defense. Lessons Learned from Ten Years of OSINT Automation. URL: https://www.youtube.com/watch?v=SMGEhFXURzY

48 SpiderFoot HX. URL: https://cybermarket.com.ua/product/spiderfoot-hx/

49 Trace Labs Blog. URL: https://www.tracelabs.org/blog

50 Trace Labs. Trace Labs OSINT VM – A Brief Tour. URL: https://www.youtube.com/watch?v=FlGdSZk1F6o&pp=ygUQVHJhY2UgTGFicyBvc2ludA%3D%3D

51 Trace Labs. Trace Labs - Open Source Intelligence Gathering. URL: https://www.youtube.com/watch?v=oz26mOwsse0&list=PL5ylEZWzbUEDsKTKdHrUVkAw0ZPqHzuj0

52 The Ultimate Beginner's Guide to OSINT. URL: https://www.osint-jobs.com/post/the-ultimate-beginners-guide-to-osint

53 The Ultimate Guide to the OSINT framework. URL: https://x-ray.contact/blog/the-ultimate-guide-to-the-osint-framework/

54 The Cyber Mentor. Open-Source Intelligence (OSINT). URL: https://www.youtube.com/watch?v=qwA6MmbeGNo

55 The OSINT Cycle: Getting familiar with the process of data collection and analysis. URL: https://osintteam.blog/the-osint-cycle-getting-familiar-with-the-process-of-data-collection-and-analysis-day3-of-6f53fcdb4234

56 Troia V. Hunting Cyber Criminals: A Hacker's Guide to Online Intelligence Gathering Tools and Techniques 1st Edition. Wiley, 2020. 544 p.

57 Udemy. URL: https://www.udemy.com/

58 Layton R., Watters P.A. Automating Open Source Intelligence: Algorithms for OSINT. Syngress. 2015. 222 p.

59 Mitnick K.D., Simon W.L. The Art of Deception: Controlling the Human Element of Security. Wiley. 2003. 368 p.

60 OSINT 2021 guide: tools and techniques for threat intelligence. URL: https://www.authentic8.com/blog/OSINT-2021-guide-tools-and-techniques

61 Open Source Intelligence (OSINT) Guide. URL: https://www.eweek.com/big-data-and-analytics/open-source-intelligence-osint/

62 A Guide to Open Source Intelligence. URL: https://www.cjr.org/tow_center_reports/guide-to-osint-and-hostile-communities.php

63 Virtual machines for OSINT. URL: https://www.learnallthethings.net

ДОДАТКИ

Додаток А

Дослідження ІнформНапалм та діяльність розвідувальних служб європейських країн

ІнформНапалм є волонтерським проєктом, призначеним для інформування про російсько-терористичну агресію проти України. Заснований в 2014 році під час початку російської збройної агресії та анексії Криму, проєкт перекладає свої матеріали на десятки мов, включаючи японську та китайську. Керівником є Роман Бурко, а речником — Михайло Макарук. ІнформНапалм зосереджується на розслідуваннях через використання відкритих джерел інформації (OSINT), зокрема соціальних мереж. Основні напрямки діяльності включають аналітику ситуації на фронті в зоні АТО, регулярні аналітичні звіти та співпрацю з інсайдерами на окупованих територіях [4].

Проєкт здійснив розслідування щодо катастрофи «Боїнга-777», створив бази даних російських військових підрозділів, а також оприлюднив інформацію про військову техніку російського походження на лінії конфлікту. Інформація видається 22 мовами, і співпраця з волонтерами з Криму, Грузії, Донбасу та інших країн світу є важливою частиною діяльності проєкту. Розслідування щодо командира 53-ї зенітно-ракетної бригади, відповідального за трагедію рейсу MH17, були використані в звіті групи Bellingcat [12], [27].

16 травня 2015 р. Роман Бурко, на своїй сторінці у Facebook, інформував про потрапляння двох російських спецпризначенців до полону українських силовиків під час бою під Щастям. Наступного дня він закликав керівництво держави та компетентні структури використовувати цей факт для удару по російській пропаганді. О 14:50, 17 травня 2015 р., лікар Григорій Максимець підтвердив інформацію про поранених російських військових та опублікував їхні фотографії на своїй сторінці у Facebook. InformNapalm висловив сподівання, що українська влада не замовчить факт взяття в полон російських спецпризначенців і використає його для контрпропаганди.

OSINT-розвідник Anton Pavlushko виявив у соціальних мережах дані про командира загону майора Костянтина Напольських, керівника групи ГРУ РФ у Луганську. Після початку російської інтервенції до Сирії, InformNapalm почав публікувати особисті дані російських пілотів та створив інфографіку разом із медіа-проєктом Visuals. Це викликало негативну реакцію в Росії, і відповідь авторів була визначена як готовність до подальших розслідувань у разі порушення режиму припинення вогню на Донбасі.

За 2016 рік спільнота здійснила 407 публікацій-розслідувань, 1197 перекладів та понад 3920 репостів у ЗМІ. У березні 2017 р., в Українському кризовому медіацентрі презентували книгу "Донбас в огні", яка включає докази російської агресії, виявлені волонтерами ІнформНапалм. В квітні 2018 р., InformNapalm опублікував інтерактивну базу даних російської агресії, що систематизує 1700 OSINT-розслідувань, включаючи російське озброєння на Донбасі та участь російських військових у конфліктах проти України, Грузії та Сирії.

У вересні 2022 р. Гродненський обласний суд засудив автора статей Дениса Івашина за ст. 365 та ч. 1 ст. 356 Кримінального кодексу, наклавши покарання у вигляді 13 років та 1 місяця умовної колонії в умовах посиленого режиму.

Спільнота ІнформНапалм активно взаємодіяла з кіберальянсом, отримуючи дані з їхніх зламів для аналізу та оприлюднення. Зокрема, отримано інформацію з телефону російського окупанта Миколи Рейхенау та розміщено відеоматеріали, викликавши реакцію YouTube.

Попередження ІнформНапалм про загрозу національній безпеці від антивіруса NOD32 та отримання даних щодо витрат Міноборони РФ на озброєння також відзначаються.

Сюжет на ARD та передача даних з розвідувального управління ЗС РФ є прикладами ефективної співпраці групи ІнформНапалм та кіберальянсу в аналізі та оприлюдненні інформації.

У квітні 2018 року ІнформНапалм, аналізуючи ексклюзивні дані від Українського кіберальянсу, оприлюднив звіт щодо участі 18-ї ОСМБр ЗС РФ у захопленні Криму. Волонтери використали виписки з наказів про харчування

військовослужбовців, відряджених на бойове завдання в окупований Крим, і виконали розвідку відкритих джерел. За результатами розслідування було ідентифіковано 40 російських військовослужбовців з 15-ї ОМСБр Збройних Сил (РФ), які здійснювали бойові операції в Луганській області та брали участь у захопленні Криму в період з 2014 року.

Технологія OSINT є важливим інструментом в галузі конкуренції, особистої, корпоративної та національної безпеки. Україна від 2014 року експериментує з її застосуванням у військових операціях, але використання цього інструменту в державному управлінні та політиці безпеки залишається на етапі досліджень.

Розслідування кіберінцидентів — трудомісткий процес, що вимагає детального аналізу. Першочергові фактори включають тип та тривалість інциденту. Збір інформації є вихідною точкою для розслідування з використанням системи OSINT.

OSINT виявився ефективним у встановленні російських військових підрозділів та незаконних бойовиків через соцмережі, а також в діяльності центру «Миротворець» та спільноти Bellingcat [6], [12]. Дослідження OSINT також визначило потенційну загрозу вторгнення Росії до України у 2022 році за допомогою Google Maps.

Крім того, OSINT виявив ремонт російської бронетехніки на підприємстві Укроборонпрому, що призвело до ракетної атаки та руйнувань. Ця подія, розголошена за допомогою OSINT, стала об'єктом пропагандистської кампанії Росії та обговоренням в українському суспільстві.

Аналіз відкритих джерел інформації проводиться у всіх країнах з розвиненими розвідувальними службами, проте спосіб організації цієї діяльності відрізняється в залежності від країни [31].

Військова доктрина **США**, визначена в "Field Manual Interim № 2-22.9", підкреслює, що в основі OSINT лежать джерело, інформація та методи їх збору, а не конкретні технічні або людські ресурси. Дослідження Hassan та Hijazi вказує на труднощі аналізу великого об'єму неструктурованої інформації, обговорює проблеми надійності джерел та використання спеціальних програмних засобів.

Williams та Blum зауважують, що лише частина отриманої OSINT інформації є релевантною, та висувають вимоги до нових методик перетворення неструктурованої інформації на звіти для політичного керівництва. Açar акцентує увагу на масовому використанні OSINT як державними службами, так і приватним сектором, звертаючи увагу на необхідність етичного та юридичного обмеження збору інформації, щоб уникнути наслідків для окремих осіб та суспільства. Одним із основних обмежень є відсутність новітніх програмних та апаратно-програмних засобів для ефективного добування, обробки та аналізу інформації в Україні, що визначається відсутністю відповідних підрозділів органів військового управління. Термін "розвідка з відкритих джерел" (OSINT) був введений розвідувальним співтовариством США в 1941 році і використовується як рівнозначний термінам "конкурентна розвідка" та "бізнес-розвідка" в інших джерелах.

В **Бельгії** відсутня загальнодержавна нормативна база для регулювання OSINT. Організаційно входить до складу Штабу Оборони Бельгії – Головної служби розвідки та безпеки (SGRS). Діяльність підрозділу OSINT визначається головним планом розвідки, що затверджується урядом. Персонал розробляє плани збору інформації, а також використовує різні джерела, включаючи FACTIVA, з бюджетом близько 650 000 євро на рік. Підрозділ, що складається з 7 осіб, розглядає понад 1500 запитів на інформацію річно та використовує різні бази даних, такі як FACTIVA та Lexis Nexis, яка була замінена через меншу продуктивність. Вибір баз даних здійснюється через тендерні процедури з урахуванням потреб підрозділу. Згідно з доповідями, близько 80% необхідної інформації можна отримати в інтернеті на платній основі.

В **Республіці Болгарія** відсутня загальнодержавна нормативна база для системи OSINT. Діяльність спецслужб РБ у цьому напрямку регулюється Законом «Про Міністерство внутрішніх справ», Законом РБ «Про спеціальні розвідувальні засоби», та іншими відомчими нормативними актами. Органи, що відповідають за добування розвідувальної інформації з відкритих джерел, входять до складу різних спецслужб РБ, таких як Національна розвідувальна служба, Національна

служба охорони, Державна агенція «Національна безпека», і Служба «Військова інформація» МО РБ. У країні розроблені та використовуються програмні продукти, зокрема DAXY Global і DAXY007, для моніторингу джерел інформації та автоматичного виявлення повідомлень за тематикою. Спецслужби США активно залучають болгарських громадян, особливо колишніх розвідників, експертів та представників ЗМІ, для здійснення роботи з відкритими джерелами. Дані свідчать, що від 65% до 85% розвідувальної інформації здобувається спецслужбами РБ із відкритих джерел [33].

У **Королівстві Великобританія** діяльність системи отримання інформації з відкритих джерел (РІВД) має глобальний характер і координується на найвищому політичному рівні. Відповідні підрозділи існують у всіх спецслужбах та численних державних установах, керівництво яких регулюється законом «Про розвідувальні служби». РІВД Великобританії використовує визначення, визначене ААР-6, та опирається на законодавчі та відомчі норми, а також на вказівки Об'єднаного розвідувального комітету Великобританії та стандарти НАТО. Система РІВД, яка функціонує в складі Міністерства оборони Великобританії, включає до 25 осіб, спеціалізованих у добуванні розвідувальної інформації з відкритих джерел. Ця система взаємодіє з розвідувальними структурами країни та має за завдання забезпечення необхідною та своєчасною інформацією інформаційно-аналітичних підрозділів ВР Великобританії. Моніторинг джерел інформації визначається пріоритетами та завданнями інформаційно-аналітичних підрозділів ВР Великобританії, які діють у складі Штабу розвідки МО.

У **Іспанії**, діяльність підрозділів розвідки з відкритих джерел регулюється законодавчою базою, такою як Закон № 23/2006, Королівський декрет № 1/1996, та Закон № 15/1999 про захист даних персонального характеру. Відповідні підрозділи працюють у всіх спецслужбах та багатьох міністерствах та відомствах. У Розвідувальному центрі Збройних сил Іспанії (ЗЦ ЗС) існує Група OSINT, яка має за завдання отримання інформації від відкритих джерел. Група працює щоденно та, за потреби, 24/7, координуючи свою роботу через органи CCIRM та діючи за трема напрямами: інформаційні агентства, зони операцій та решта

інформації. Технічні засоби, такі як програми "ad-hoc" та "Hardware", використовуються для анонімного доступу до Інтернету, пошуку та здобуття інформації, а також трансформації аудіо- та відеосигналів у текст. Пошук розділений на щоденний та щомісячний цикли, останній здійснюється згідно з завданнями PROGINT та за пріоритетами визначеними для розвідки. Також може проводитися пошук за запитами від CCIRM.

В **Італії** діяльність розвідувальних органів та спецслужб, спрямована на збір та обробку інформації з відкритих джерел, регулюється законом № 124 від 3 серпня 2007 року "Система розвідки та безпеки Італійської Республіки та новий порядок забезпечення державної таємниці". Згідно із Законом № 133 від 7 серпня 2012 року, який вніс зміни до зазначеного закону, визначено нові повноваження та завдання парламентського комітету з питань безпеки та визначено функції спецслужб.

Інформація щодо структури підрозділів OSINT італійських секретних служб є конфіденційною, оскільки її розкриття може викликати ризик розголошення оперативних аспектів. У Збройних силах Італії, Управління інформації та безпеки (УІБ) Генерального штабу веде розвідувально-інформаційну діяльність за кордоном. Управління не має єдиного інформаційно-аналітичного підрозділу, тому кожен його підрозділ готує матеріали відповідно до свого напряму [34].

Міжвидовий розвідувальний центр, що входить до складу УІБ, координує інформаційне забезпечення операцій за межами країни та співпрацює із Службою розвідки управління операцій Вищого міжвидового оперативного командування. Останній має підпорядковані розвідувальні відділи видів ЗС Італії та здійснює збір, обробку та передачу розвідувальної інформації, включаючи дані з відкритих джерел.

У цивільних органах, Департамент розвідки та безпеки при Раді Міністрів Італії, координуючи роботу секретних служб, забезпечує підготовку інформаційно-аналітичних документів та надає рекомендації вищому державному керівництву на основі розвідувальної інформації від інформаційних підрозділів Агентства розвідки та зовнішньої безпеки, а також спеціальних органів Військ

карабінерів, Фінансової гвардії, Державної поліції та ЗС Італії. При підготовці доповідей та аналітичних матеріалів, спецслужби Італії використовують до 60% інформації з відкритих джерел для перевірки достовірності, порівняння та підтвердження їх змісту [35].

В **Республіці Польща** відсутня окрема нормативна база для діяльності спецслужб та інституцій по роботі з відкритими джерелами інформації (OSINT). Кожна служба впроваджує цю діяльність в межах чинного законодавства та своєї компетенції.

Хоча відсутні окремі структури для управління OSINT в розвідувальних структурах Польщі, розвідувальні підрозділи на всіх рівнях використовують цей метод з урахуванням їхнього технічного оснащення. Основною особливістю системи інформаційно-аналітичного забезпечення є використання органами стратегічного аналізу та прогнозування військово-політичної обстановки розвинутої системи отримання та обробки інформації.

Розвідувальні органи використовують відкриті джерела інформації з різних джерел, включаючи внутрішні підрозділи, міжнародні організації, комерційні центри та аналогічні інституції країн-союзників. Наприклад, Міністерство закордонних справ РП користується послугами The Economist Intelligence Unit для альтернативного джерела інформації та прогнозування розвитку ситуації в світі.

Для ефективного data mining українські розвідувальні органи використовують програмне забезпечення Chost Miner (3.0), розроблене Кафедрою інформатики Університету в Торуню та комерціалізоване фірмою FQS Poland. Це програмне забезпечення включає інструмент для створення моделі знань та інструмент для аналізу, відображення та використання даних.

Федеративна Республіка Німеччина реалізує таємні програми, спрямовані на розбудову здатностей правоохоронних та спеціальних служб для добування та аналізу інформації у відкритих інформаційних системах в інтересах національної безпеки.

Федеральна розвідувальна служба (Bundesnachrichtendienst) впроваджує Концепцію "Стратегічна технологічна ініціатива", що включає 26 проектів і

передбачає розширення апаратного та програмного забезпечення для вдосконалення електронної розвідки, зокрема, моніторингу соціальних мереж. Розробка техніко-економічного обґрунтування почалася у 2013 році.

Федеральне міністерство оборони ФРН реалізує проект "Отримання інформації з відкритих джерел" (Wissenserschließung aus offenen Quellen, WeroQ), спрямований на створення системи автоматизованого аналізу відкритих джерел інформації для вирішення завдань воєнної безпеки. Головний підрядник – Науково-дослідний інститут зв’язку, обробки інформації та ергономіки (Forschungsinstitut für Kommunikation, Informationsverarbeitung und Ergonomie, FhG FKIE), субпідрядник – компанія "IBM" (США).

Додаток Б

Код для підключення до СУБД

```
class Neo4jConnection:
    def __init__(self, uri, user, password):
        self.driver = GraphDatabase.driver(uri, auth=(user, password))

    def close(self):
        if self.driver is not None:
            self.driver.close()

    def query(self, query, db=None):
        assert self.driver is not None, "Driver not initialized!"
        session = None
        response = None
        try:
            session = self.driver.session(database=db) if db is not None else self.driver.session()
            response = list(session.run(query))
        except Exception as e:
            print("Query failed:", e)
        finally:
            if session is not None:
                session.close()
        return response
```

Додаток В

Код системи пошуку

```
curl -XGET "$ES_URL/blog/post/_search?pretty" -d'
{
  "_source": false,
  "query": {
    "match": {
      "content": " story "
    }
  }
}'

{
  "took" : 13,
  "timed_out" : false,
  "_shards" : {
    "total" : 5,
    "successful" : 5,
    "failed" : 0
  },
  "hits" : {
    "total" : 3,
    "max_score" : 0.11506981,
    "hits" : [ {
      "_index" : "blog",
      "_type" : "post",
      "_id" : "2",
      "_score" : 0.11506981
    }, {
      "_index" : "blog",
      "_type" : "post",
      "_id" : "1",
      "_score" : 0.11506981
    }, {
      "_index" : "blog",
      "_type" : "post",
      "_id" : "3",
      "_score" : 0.095891505
    } ]
  }
}
```

```
curl -XGET "$ES_URL/_analyze?pretty&analyzer=standard&text=%D0%92%D0%B5%D1%81%D0%B5%D0%BB%D1%8B%D0%B5%20%D0%B8%D1%81%D1%82%D0%BE%D1%80%D0%B8%D0%B8%20%D0%BF%D1%80%D0%BE%20%D0%BA%D0%BE%D1%82%D1%8F%D1%82"

{
  "tokens" : [ {
    "token" : "stories",
    "start_offset" : 0,
    "end_offset" : 7,
    "type" : "<ALPHANUM>",
    "position" : 0
  }, {
    "token" : "love",
    "start_offset" : 8,
    "end_offset" : 15,
    "type" : "<ALPHANUM>",
    "position" : 1
  }, {
    "token" : " about",
    "start_offset" : 16,
    "end_offset" : 19,
    "type" : "<ALPHANUM>",
    "position" : 2
  }, {
    "token" : "war",
    "start_offset" : 20,
    "end_offset" : 25,
    "type" : "<ALPHANUM>",
    "position" : 3
  } ]
}
curl -XPOST "$ES_URL/blog2" -d'
{
  "settings": {
    "analysis": {
      "filter": {
        "ua_stop": {
          "type": "stop",
          "stopwords": "_ Ukrainian _"
        },
        "ua_stemmer": {
          "type": "stemmer",
```

```
        "language": " Ukrainian"
       }
      },
      "analyzer": {
       "default": {
        "char_filter": [
         "html_strip"
        ],
        "tokenizer": "standard",
        "filter": [
         "lowercase",
         "ua_stop",
         "ua_stemmer"
        ]
       }
      }
    },
    "mappings": {
     "post": {
      "properties": {
       "content": {
        "type": "string"
       },
       "published_at": {
        "type": "date"
       },
       "tags": {
        "type": "string",
        "index": "not_analyzed"
       },
       "title": {
        "type": "string"
       }
      }
     }
    }
}'
curl -XPOST "$ES_URL/blog2/post/_search?pretty" -d'
{
  "query": {
    "simple_query_string": {
      "query": "stories",
      "fields": [
```

```
        "title^3",
        "tags^2",
        "content"
      ]
    }
  }
}'
curl -XPOST "$ES_URL/blog2/post/_search?pretty" -d'
{
  "query": {
    "simple_query_string": {
      "query": "- love",
      "fields": [
        "title^3",
        "tags^2",
        "content"
      ]
    }
  }
}'
```

Додаток Г

Встановлення та налаштування

```
dbms.unmanaged_extension_classes=com.graphaware.server=/graphaware
com.graphaware.runtime.enabled=true
com.graphaware.module.UIDM.1=com.graphaware.module.uuid.UuidBootstrapper
com.graphaware.module.UIDM.uuidProperty=uuid
com.graphaware.module.UIDM.node=hasLabel('Label1') || hasLabel('Label2')
com.graphaware.module.UIDM.uuidIndex=uuidIndex
com.graphaware.module.UIDM.initializeUntil=0
com.graphaware.module.ES.2=com.graphaware.module.es.ElasticSearchModuleBootstr
apper
com.graphaware.module.ES.uri=localhost
com.graphaware.module.ES.port=9201
com.graphaware.module.ES.protocol=http
com.graphaware.module.ES.index=neo4j-index
com.graphaware.module.ES.keyProperty=uuid
com.graphaware.module.ES.retryOnError=false
com.graphaware.module.ES.queueSize=10000
com.graphaware.module.ES.reindexBatchSize=2000
com.graphaware.module.ES.node=hasLabel('Person')
com.graphaware.module.ES.node.property=key != 'age'
com.graphaware.module.ES.bulk=true
com.graphaware.module.ES.initializeUntil=0
GraphAwareRuntime runtime = GraphAwareRuntimeFactory.createRuntime(database);
//where database is an instance of GraphDatabaseService
runtime.registerModule(new UuidModule("UUID",
UuidConfiguration.defaultConfiguration(), database));
configuration = ElasticSearchConfiguration.defaultConfiguration(HOST, PORT);
runtime.registerModule(new ElasticSearchModule("ES", new
ElasticSearchWriter(configuration), configuration));

runtime.start();
```

Додаток Д

Код програми

Mainwindow.h

```
#ifndef MAINWINDOW_H #define MAINWINDOW_H

#include <QUrl>
#include <QMainWindow>
#include <QtNetwork/QNetworkAccessManager> #include
<QtNetwork/QNetworkAccessManager> #include <QtNetwork/QNetworkRequest> #include
<QtNetwork/QNetworkReply>
#include <QProcess> #include <QMessageBox> #include <QKeyEvent>

#include <diagrammainwindow.h>


QT_BEGIN_NAMESPACE
namespace Ui { class MainWindow; }
QT_END_NAMESPACE


class MainWindow : public QMainWindow
{
   Q_OBJECT
public:
   MainWindow(QWidget *parent = nullptr);
   ~MainWindow();

public slots:
   void startSearch();
   void specifyUrlPessed();
   void specifyFilePressed();
   void clearGroupPressed();
   void replyFinished(QNetworkReply *reply);
   void searchReplyFinished(QNetworkReply *reply); void itemSelected(const QModelIndex
   &index);
   void pythonFinished(int exitCode, QProcess::ExitStatus exitStatus); void startScript();
protected:
   void keyPressEvent(QKeyEvent *event);

private:
   DiagramMainWindow *mDiagramWIndow; QMessageBox *mMessageBox;
   QDialog *idleDialog; QProcess *mProcess;
   QNetworkAccessManager *mManager; Ui::MainWindow *ui;
};
#endif // MAINWINDOW_H
```

Mainwindow.cpp

```
#include "mainwindow.h" #include "ui_mainwindow.h"
```

```
#include <libxml2/libxml/HTMLparser.h> #include <libxml2/libxml/parser.h> #include
<libxml2/libxml/tree.h>

#include <QInputDialog> #include <QJsonDocument> #include <QJsonObject> #include
<QJsonArray> #include <QDesktopServices>
#include <QUrl>

#define BORDER 600

MainWindow::MainWindow(QWidget *parent)
    : QMainWindow(parent)
    , mDiagramWIndow(new DiagramMainWindow())
    , ui(new Ui::MainWindow)
    , mMessageBox(new QMessageBox(this))
    , idleDialog(new QDialog(this))
    , mManager(new QNetworkAccessManager(this))
    , mProcess(new QProcess(this))
{
    ui->setupUi(this);

    mMessageBox->setModal(true);
    mMessageBox->setWindowTitle("Wait please!"); mMessageBox->setText("
                                                                        Idle.
    ..                                                                  ");
    mMessageBox->setStandardButtons(0); mMessageBox->setMinimumHeight(300);
    mMessageBox->setMinimumHeight(400);

    connect(mManager, &QNetworkAccessManager::finished, this,
&MainWindow::replyFinished);

    connect(mProcess, static_cast<void(QProcess::*)(int,
QProcess::ExitStatus)>(&QProcess::finished), this, &MainWindow::pythonFinished);

    connect(ui->specifyUrlPushButton, &QPushButton::pressed, this,
&MainWindow::specifyUrlPressed);
    connect(ui->specifyFilePushButton, &QPushButton::pressed, this,
&MainWindow::specifyFilePressed);
    connect(ui->lineEdit, &QLineEdit::returnPressed, this, &MainWindow::startSearch);
    connect(ui->resultListWidget, &QListWidget::clicked, this, &MainWindow::itemSelected);
    connect(mDiagramWIndow, &DiagramMainWindow::startScript, this,
&MainWindow::startScript);

    ui->resultListWidget->setSpacing(4);

    ui->resultListWidget->setSpacing(4);
}

MainWindow::~MainWindow()
{
```

```
    delete ui;
}

void MainWindow::startSearch()
{
    QUrl serviceUrl = QUrl("http://localhost:9200/_search?pretty"); auto requestText = ui-
    >lineEdit->text();
    QByteArray postData;
    //User request to search system postData.append("{\
                \"_source\": [\
                  \"title\",\
                  \"content\"\
                ],\
                \"query\": {\
                  \"match\": {\
                    \"content\": \"" + requestText + "\"\
                  }\
                }\
                                                    }");
                                            // Call the webservice
    QNetworkAccessManager *manager = new QNetworkAccessManager(this);

    QNetworkRequest request(serviceUrl);
    request.setHeader(QNetworkRequest::ContentTypeHeader, QVariant(QString("text/xml")));

   connect(manager, &QNetworkAccessManager::finished, this,
&MainWindow::searchReplyFinished);
   manager->post(request, postData);
}

void MainWindow::specifyUrlPressed()
{
    bool ok;
    QString text = QInputDialog::getText(0, "Input dialog","Specify Url:", QLineEdit::Normal, "",
    &ok);
    if (ok && !text.isEmpty())
    {
        qDebug() << "specified URL " << text;
        QStringList arguments { "/home/max/InformationGatheringTool/sendDataScript.py",
QString("-url=%1").arg(text) };
        mProcess->start("python3", arguments); mMessageBox->show();
        mProcess->waitForFinished();
    }
}

void MainWindow::specifyFilePressed()
{
    mDiagramWIndow->show();
}
```

```
void MainWindow::clearGroupPressed()
{

}

void MainWindow::replyFinished(QNetworkReply *reply)
{
   QByteArray buffer = reply->readAll();

   qDebug()<<"Reply received! "<< buffer;
}

void MainWindow::searchReplyFinished(QNetworkReply *reply)
{
   QByteArray buffer = reply->readAll(); ui->resultListWidget->clear(); if (buffer.size())
   {
      QString title; QString content; double score= 0.0;

      QJsonDocument document = QJsonDocument::fromJson(buffer);


      QJsonObject root = document.object();


      QJsonValue hitsExternalVal = root.value("hits"); if (hitsExternalVal.isObject())
      {
         QJsonObject hitsExternalObj = hitsExternalVal.toObject(); QJsonValue hitsVal =
         hitsExternalObj.value("hits");
         if(hitsVal.isArray()){
            QJsonArray hitsArray = hitsVal.toArray();

            for(int i = 0; i < hitsArray.count(); i++){  QJsonObject subtree =
               hitsArray.at(i).toObject(); score = subtree.value("_score").toDouble(); QJsonValue
               sourceVal = subtree.value("_source"); if (sourceVal.isObject())
               {
                  QJsonObject sourceObj = sourceVal.toObject(); title =
                  sourceObj.value("title").toString(); content =
                  sourceObj.value("content").toString(); auto simplifiedString =
                  content.simplified();
                  int index = simplifiedString.indexOf(ui->lineEdit->text()); QString stringToShow =
                  simplifiedString.left(BORDER); bool paintText = false;
                  if (index > 0)
                  {
                     paintText = true;
                     int count = index + BORDER < simplifiedString.size() ? BORDER :
simplifiedString.size() - index;
                     stringToShow = simplifiedString.mid(index, count);
                  }
                  //result element window
```

```
                QWidget* window = new QWidget();


                QLabel *label = new QLabel();
                label->setText(QString("<b>%1<b>").arg(title)); label-
                >setAlignment(Qt::AlignLeft);

                QLabel *label2 = new QLabel(); if (paintText)
                {
                   label2->setText(QString("<fontcolor=\"red\">%1</font>").arg(stringToShow));
                }
                else
                {
                   label2->setText(QString("%1").arg(stringToShow));
                }
                label2->setAlignment(Qt::AlignLeft);

                QListWidgetItem* item;
                item = new QListWidgetItem(ui->resultListWidget); item-
                >setSizeHint(QSize(200,80)); // you could change it

                //Creating layout for our label and etc... QVBoxLayout* layout = new
                QVBoxLayout();
                //Adding elements to layout layout->addWidget(label); layout-
                >addWidget(label2);

                //layout->addStretch(); window->setLayout(layout);

                //Adding the item to the listwidget item->setData(Qt::UserRole ,title);

                ui->resultListWidget->addItem(item);
                item->setToolTip(QString("score = ") + QString::number(score)); ui-
                >resultListWidget->setItemWidget(item,window);

                QString widgetText;
                QString fileName = "fileName";
                widgetText= fileName +" "+ "Date";
             }
          }
       }
    }
 }

 qDebug()<<"Reply received! "<< buffer;
}

void MainWindow::itemSelected(const QModelIndex &index)
{
```

```
    QString str = index.data(Qt::UserRole).toString(); QString url = str.split(" ").at(0);
    url.replace("website", ""); QDesktopServices::openUrl(QUrl(url)); qDebug()<< url;
}

void MainWindow::pythonFinished(int exitCode, QProcess::ExitStatus exitStatus)
{
    qDebug()<< "Python script finished with code "<< exitCode; mMessageBox->hide();
}

void MainWindow::startScript()
{
    QStringList arguments { "/home/max/InformationGatheringTool/sendFiledataScript.py"};
    mProcess->startDetached("python3", arguments);
    mMessageBox->show();
    // mProcess->waitForFinished();
}

void MainWindow::keyPressEvent(QKeyEvent *event)
{
    switch(event->key()) {
      case Qt::Key_Space: { qDebug() << "Space";  break; }
      case Qt::Key_Left: { qDebug() << "Left"; mMessageBox->hide(); break; } case Qt::Key_Right:
      { qDebug() << "Right"; break; }
      default: { qDebug() << "Unhandled"; break; }
      }
}
```

sendDataScript.py

```
from py2neo import Graph import requests
from bs4 import BeautifulSoup from bs4.element import Comment import argparse
import sys

def tag_visible(element):
    if element.parent.name in ['style', 'script', 'head', 'title', 'meta', '[document]']: return False
    if isinstance(element, Comment): return False
    return True

def text_from_html(page):
    soup = BeautifulSoup(page.content, 'html.parser') texts = soup.findAll(text=True)
    visible_texts = filter(tag_visible, texts)
    return u" ".join(t.strip() for t in visible_texts)

def remove_copy(links): seen = set()
    result = []
    for item in links:
        if item not in seen: seen.add(item) result.append(item)
    return result

if __name___== '__main__':
```

```
        parser = argparse.ArgumentParser()

        parser.add_argument('-url', action='store', dest='URL', help='specify URL to scrap')
        arguments = parser.parse_args()

        if arguments.URL:
            URL = arguments.URL else:
            parser.print_help() sys.exit(1)

page = requests.get(URL)

        soup = BeautifulSoup(page.content, 'html.parser') raw_links = soup.find_all('a')

        link_set = set() link_set.add(URL)
        for tag in raw_links:
            link = tag.get('href', None)
            if link is not None and link.startswith('https'): link_set.add(link)

        links = remove_copy(link_set)

        for link in link_set: print(link)
            page = requests.get(link)
            cleanText = text_from_html(page).replace("'", ".").replace("$", " ").replace("\\", " ").replace("//", " ")
            graph = Graph("http://localhost:7474", auth=("neo4j", "neo4j"))
            result = graph.run("CREATE (website:Website { title:'website%s', content:'%s' })" % (link, cleanText))
            print('create result ', result) for link in link_set:
            graph = Graph("http://localhost:7474", auth=("neo4j", "klkl"))
            result = graph.run("MATCH (a:Website), (b:Website) WHERE a.title = 'website%s' AND b.title = 'website%s' CREATE(a) - [: MENTIONED]->(b)" % (URL, link))
            print('create relationship ', result) print(len(link_set)
```