МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра кібербезпеки

РОМАНЮК Орест-Остап Мар'янович

Моніторинг темних веб-ресурсів за допомогою відкритого джерела інформації для запобігання кіберзлочинності / Monitoring Dark Web Resources Using Open Source Intelligence to Prevent Cybercrime

спеціальність: 125 – Кібербезпека та захист інформації

освітньо-професійна програма – Кібербезпека

Кваліфікаційна робота

Виконав студент групи
КБм -21
О. О. М. Романюк

Науковий керівник
к.т.н., доцент Цаволик Т.Г.

Кваліфікаційну роботу
Допущено до захисту:

«____» _____ ___2024 р.

Завідувач кафедри
_____ **В.В.Яцків**

ТЕРНОПІЛЬ – 2024

Факультет комп'ютерних інформаційних технологій
Кафедра кібербезпеки
Освітній ступінь «магістр»
спеціальність: 125 – Кібербезпека та захист інформації освітньо-професійна програма – Кібербезпека

ЗАТВЕРДЖУЮ
Завідувач кафедри

_____ В.В.Яцків
«____» _____ 2023 року

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

РОМАНЮК Орест-Остап Мар'янович
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи:
Моніторинг темних веб-ресурсів за допомогою відкритого джерела інформації для запобігання кіберзлочинності / Monitoring Dark Web Resources Using Open Source Intelligence to Prevent Cybercrime

керівник роботи к.т.н., доцент Цаволик Т. Г.
затверджені наказом по університету від 12 грудня 2023 року № 753
2. Строк подання студентом закінченої кваліфікаційної роботи 12 грудня 2024 р.

3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.

4. Основні питання, які потрібно розробити:
– Визначити ключові особливості та структуру Dark Web.
–Класифікувати типи відкритих джерел, релевантних для моніторингу Dark Web.
–Проаналізувати основні типи кіберзлочинів, що здійснюються через Dark Web.
– Дослідити забезпечення анонімності для .onion
– Спроектувати загальну архітектуру системи моніторингу Dark Web.

5. Перелік графічного матеріалу у роботі:
-- Схема роботи VPN
– Схема алгоритму з'єднання Tor браузера з кінцевим хостом.
– Схема взаємодії модулів в Connection_pool.
– Схема алгоритму роботи проксі – сервера SOCKS5.

– Схема роботи алгоритму парсингу з CSV

6. Консультанти розділів кваліфікаційної роботи

	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання 8 грудня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строки виконання етапів кваліфікаційної роботи	Примітка
1	Аналіз предметної області	12.2023 р. – 03.2024 р.	
2	Техніки дослідження моніторингу Dark Web.	03.2024 р. – 06.2024 р.	
3	Проектування та розробка засобу моніторингу Dark Web.	06.2024 р. – 11.2024 р.	

Студент _____ Романюк О.О.М.
(підпис)

Керівник роботи _____ к.т.н., доцент Цаволик Т.Г.
(підпис)

АНОТАЦІЯ

Кваліфікаційна робота на тему «Моніторинг темних веб-ресурсів за допомогою відкритого джерела інформації для запобігання кіберзлочинності» на здобуття освітнього ступеня «Магістр» зі спеціальності 125 «Кібербезпека та захист інформації» освітньо-професійної програми «Кібербезпека» написана обсягом 95 сторінок і містить 33 ілюстрацій, 7 таблиць, 2 додатки та 58 джерел за переліком посилань.

Метою кваліфікаційної роботи є розробка та демонстрація програмної реалізації моніторингу темних веб-ресурсів з використанням відкритих джерел для запобігання кіберзлочинності. Дослідження спрямоване на побудову комплексного підходу для підвищення ефективності виявлення, аналізу та протидії кіберзагрозам, що виникають у середовищі Dark Web.

Створено систему моніторингу дарквеб-ресурсів з використанням технологій автоматизованого підключення до мережі Tor. Розроблено модулі для автоматичного встановлення з'єднання, зміни IP-адрес та управління налаштуваннями проксі.

Створено модулі для обробки динамічного контенту та асинхронного завантаження даних. Реалізовано алгоритми фільтрації та категоризації отриманої інформації для подальшого аналізу.

Розроблено механізми формування та обробки CSV-файлів для зберігання зібраних даних. Реалізовано алгоритми очищення та нормалізації даних, видалення дублікатів та форматування інформації відповідно до визначеної структури.

Ключові слова: OSINT, Dark Web, ONION.

ABSTRACT

The qualification work on the topic "Monitoring dark web resources using open source information to prevent cybercrime" for obtaining the degree of "Master" in specialty 125 "Cybersecurity and information protection" of the educational and professional program "Cybersecurity" is written in 95 pages and contains 33 illustrations, 7 tables, 2 appendices and 58 sources according to the list of references.

The purpose of the qualification work is to develop and demonstrate the software implementation of monitoring dark web resources using open sources to prevent cybercrime. The research is aimed at building a comprehensive approach to increase the efficiency of detecting, analyzing and countering cyber threats that arise in the Dark Web environment.

A system for monitoring dark web resources has been created using technologies for automated connection to the Tor network. Modules have been developed for automatic connection establishment, changing IP addresses and managing proxy settings.

Modules for processing dynamic content and asynchronous data loading have been created. Algorithms for filtering and categorization of the received information for further analysis have been implemented.

Mechanisms for generating and processing CSV files for storing the collected data have been developed. Algorithms for cleaning and normalizing data, removing duplicates, and formatting information according to a defined structure have been implemented.

Keywords: OSINT, Dark Web, ONION.

ЗМІСТ

СПИСОК ВИКОРИСТАНИХ СКОРОЧЕНЬ

OSINT - Open Source Intelligence.

ПЗ - Програмне забезпечення.

I2P - Invisible Internet Project.

ЗМІ - Засоби масової інформації.

XML - Extensible Markup Language.

HTML - HyperText Markup Language.

CSS - Cascading Style Sheets.

CSV - Comma-Separated Values.

IP - Internet Protocol.

VPN - Virtual Private Network.

DOM - Document Object Model.

GUI - Graphical User Interface.

UI - User Interface.

HTTP - HyperText Transfer Protocol.

HTTPS - HyperText Transfer Protocol Secure.

ВСТУП

Актуальність теми: Актуальність теми дослідження зумовлена стрімким розвитком інформаційних технологій та зростанням кіберзлочинності з використанням ресурсів "темної павутини". Завдяки своїй анонімності та складності відстеження, "темна мережа" стала привабливим середовищем для незаконної діяльності, включаючи торгівлю наркотиками, зброєю, персональними даними та інші форми кіберзлочинності. Традиційні правоохоронні заходи не є повною мірою ефективними у протидії цим загрозам через технічні характеристики ресурсів темного Інтернету та анонімність їхніх користувачів. Тому існує нагальна потреба у розробці та впровадженні інноваційних підходів до виявлення, моніторингу та запобігання кіберзлочинам, що вчиняються через ці канали.

Перспективним напрямком кібербезпеки є використання відкритих джерел для моніторингу темних веб-ресурсів. Такий підхід дозволяє аналізувати великі обсяги даних з легальних джерел, які можуть містити непрямі докази активності в Dark web. Інтеграція аналітичних методів з відкритих джерел та методів моніторингу темного Інтернету створює потужний інструментарій для виявлення потенційних загроз на ранніх стадіях їх формування. Такий комплексний підхід дозволить не лише боротися з кіберзлочинами, які вже були здійснені, але й прогнозувати та запобігати майбутнім атакам.

Про важливість цієї теми свідчить також зростаючий інтерес наукової спільноти та фахівців-практиків з кібербезпеки до розробки методологій та інструментів ефективного моніторингу ресурсів інтернету. Дослідження в цій галузі спрямовані на подолання технічних, правових та етичних проблем, пов'язаних з доступом та аналізом інформації в Dark Web. Розвиток технологій машинного навчання та штучного інтелекту відкриває нові можливості для автоматизації збору та обробки даних з відкритих джерел, що може значно підвищити ефективність виявлення кіберзагроз. Водночас важливим залишається дотримання балансу між необхідністю спостереження для запобігання злочинам та захистом прав на приватність користувачів мережі.

Мета і завдання дослідження. Метою дослідження є розробка та демонстрація програмної реалізації моніторингу темних веб-ресурсів з використанням відкритих джерел для запобігання кіберзлочинності. Дослідження спрямоване на побудову комплексного підходу для підвищення ефективності виявлення, аналізу та протидії кіберзагрозам, що виникають у середовищі Dark Web.

Зазначена мета передбачає вирішення таких завдань:

- Визначити ключові особливості та структуру Dark Web.

- Класифікувати типи відкритих джерел, релевантних для моніторингу Dark Web.

- Проаналізувати основні типи кіберзлочинів, що здійснюються через Dark Web.

- Дослідити законодавчу базу щодо Dark Web.

- Дослідити забезпечення анонімності для .onion

- Проаналізувати методологію аналізу зібраних даних

- Дослідити існуючі системи скрапінгу та їх ефективність

- Спроектувати загальну архітектуру системи моніторингу Dark Web.

- Розробити програмні модулі для збору даних з відкритих джерел

- Провести комплексне тестування розробленої системи моніторингу

Науково-теоретичною основою дослідження стали праці: Гарет Оуена, Даніеля Мура, Томаса Ріда, Алекса Бірюкова, Майкла Чертоффа, Джейма Бартлета та інших.

Об'єктом дослідження є процеси виявлення та запобігання кіберзлочинності в середовищі темного інтернету.

Предметом дослідження є методи та засоби моніторингу темних веб-ресурсів з використанням відкритих джерел інформації.

Методи дослідження: методи аналізу відкритих джерел інформації (OSINT), data mining, контент-аналіз, статистичний аналіз.

Наукова новизна: Наукова новизна дослідження полягає у розробці та обґрунтуванні комплексного підходу до моніторингу темних веб-ресурсів з використанням відкритих джерел інформації для запобігання кіберзлочинності.

Запропоновано новий скрипт для етичного та легального проведення моніторингу темних веб-ресурсів з використанням відкритих джерел інформації, що враховує сучасні правові норми та етичні стандарти в сфері кібербезпеки.

Практичне значення: Практичне значення отриманих результатів дослідження полягає у створенні ефективного інструментарію для моніторингу темних веб-ресурсів з використанням відкритих джерел інформації, що має суттєвий потенціал для застосування у сфері кібербезпеки та протидії кіберзлочинності. Розроблена методологія та алгоритми можуть бути безпосередньо впроваджені у практичну діяльність правоохоронних органів, служб кібербезпеки та аналітичних центрів, що займаються виявленням та запобіганням кіберзагроз.

Результати дослідження: Результати даного дослідження сприяють розробці методів та інструментів моніторингу темних веб-ресурсів з використанням відкритих джерел з метою запобігання кіберзлочинності. В ході дослідження було отримано низку важливих наукових і практичних результатів, які мають значення для підвищення ефективності боротьби з кіберзлочинністю.

Публікації та апробація до магістерської роботи.

1. О-О. РОМАНЮК. ПРАВОВІ ТА ЕТИЧНІ АСПЕКТИ МОНІТОРИНГУ ТЕМНОГО ВЕБУ. Збірник матеріалів науково-практичного симпозіуму «Захист інформації», Тернопіль, 2024. С. 48-50

2. О-О. РОМАНЮК. ІНСТРУМЕНТИ ТА МЕТОДИ МОНІТОРИНГУ ВІДКРИТИХ ДЖЕРЕЛ. Збірник матеріалів науково-практичної конференції молодих вчених, аспірантів та студентів «Кібербезпека та комп’ютерно-інтегровані технології» (КБКІТ - 2024), Тернопіль, 2024. С. 140-142

1 ТЕОРЕТИЧНІ ОСНОВИ МОНІТОРИНГУ ТЕМНИХ ВЕБ-РЕСУРСІВ

1.1. Поняття та характеристики Dark Web.

Dark Web представляє собою частину глобальної мережі Інтернет, яка функціонує на основі особливих протоколів та технологій, що забезпечують високий рівень анонімності та приватності користувачів. Ця мережа характеризується використанням шифрування та маршрутизації трафіку через декілька проміжних вузлів, що значно ускладнює відстеження активності користувачів та їх ідентифікацію.

Концепція Dark Web є предметом інтенсивних наукових дискусій та досліджень, що призвело до формування різноманітних визначень цього феномену. У науковій літературі можна зустріти ряд підходів до трактування поняття Dark Web, кожен з яких акцентує увагу на певних аспектах цього складного явища. Характерною особливістю Dark Web є наявність так званих "прихованих сервісів" або "цибулевих сайтів", які доступні лише через спеціалізоване програмне забезпечення. Ці ресурси використовують доменне ім'я .onion та не індексуються традиційними пошуковими системами. Ніколас Крісті з Університету Карнегі-Меллона у своєму дослідженні економіки Dark Web-ринків відзначає: Приховані сервіси Tor забезпечують платформу для створення анонімних та децентралізованих ринків, які часто використовуються для нелегальної торгівлі [2].

Dark Web не є однорідним середовищем і використовується для різноманітних цілей. З одного боку, він є середовищем дисидентів, журналістів та активістів у країнах з репресивними режимами, дозволяючи їм безпечно комунікувати та поширювати інформацію. З іншого боку, анонімність Dark Web приваблює кримінальні елементи, які використовують цю мережу для незаконної діяльності. Даніель Мур та Томас Рід у своєму аналізі контенту прихованих сервісів Tor зазначають: Хоча більшість сайтів у мережі Tor пов'язані з легальною діяльністю, значна частина трафіку спрямована на ресурси, пов'язані з незаконними товарами та послугами [3].

Технічні аспекти функціонування Dark Web постійно еволюціонують, що створює нові виклики для правоохоронних органів та дослідників кібербезпеки. Алекс Бірюков, професор Люксембурзького університету, у своїй роботі з деанонімізації користувачів Tor підкреслює: "Постійний розвиток технологій анонімізації та шифрування в Dark Web вимагає розробки нових методів аналізу та моніторингу для ефективної протидії кіберзлочинності" [4] Соціальний вплив Dark Weby є предметом активних дискусій та досліджень. Джеймі Бартлетт, автор книги "The Dark Net", зазначає: "Dark Web є не лише технологічним феноменом, але й соціальним простором, де формуються унікальні спільноти та субкультури, часто на межі легальності та етики" [5]. Це підкреслює необхідність комплексного підходу до вивчення та розуміння Dark Weby, який враховував би не лише технічні, але й соціальні та психологічні аспекти цього явища.

Дослідник Джудіт Олдрідж у своїй фундаментальній праці визначають Dark Web як сукупність мереж та технологій, що використовуються для обміну цифровим контентом анонімно [6]. Це визначення підкреслює технологічний аспект Dark Weby, фокусуючись на його функціональному призначенні як засобу анонімної комунікації. Розширюючи це поняття, Ніколас Крісті пропонує розглядати Dark Web як екосистему прихованих сервісів, доступних через спеціалізовані мережі, такі як Tor, що забезпечують анонімність та шифрування комунікацій [7]. Таке трактування акцентує увагу не лише на технологічних аспектах, але й на структурних особливостях Dark Web'у як середовища функціонування специфічних онлайн-ресурсів. Федеріко Боргоново та Алі Фішер, у свою чергу, наголошують на багатошаровості поняття Dark Weby, визначаючи його як комплексну систему, що включає в себе технології анонімізації, криптографічні протоколи та децентралізовані мережеві архітектури, спрямовані на забезпечення максимальної приватності та захисту від цензури [8]. Це визначення підкреслює технічну складність та багатогранність Dark Web'у як об'єкта дослідження.

У контексті кібербезпеки Майкл Чертофф пропонує розглядати Dark Web як сегмент кіберпростору, що характеризується високим рівнем анонімності та шифрування, який може бути використаний як для легітимних, так і для

незаконних цілей [9]. Таке визначення враховує дуальну природу Dark Web'у та його потенційні наслідки для суспільства та національної безпеки.

Джеймі Бартлетт у своїх дослідженнях розширює поняття Dark Web'у, включаючи до нього соціокультурний аспект: Dark Web - це не лише технологічна інфраструктура, але й соціальний простір, де формуються специфічні онлайн-спільноти, субкультури та економічні відносини, часто на межі легальності [10]. Це визначення підкреслює важливість розгляду Dark Web'у не лише як технологічного феномену, але й як соціального конструкту.

Алекс Бірюков та його колеги пропонують технічно орієнтоване визначення, розглядаючи Dark Web як набір розподілених мережевих технологій, що забезпечують анонімність через багатошарову маршрутизацію та криптографічні протоколи [11]. Це визначення фокусується на технічних механізмах, що лежать в основі функціонування Dark Weby.

Ерік Джардін розширює поняття Dark Web'у, включаючи до нього аспект доступності інформації: Dark Web можна розуміти як частину глибокого вебу, яка навмисно прихована та доступна лише через спеціалізоване програмне забезпечення, що забезпечує анонімність користувачів [12]. Це визначення підкреслює зв'язок між Dark Web'ом та іншими концепціями, такими як глибокий та прихований веб. Роджер Діндледайн, один з розробників мережі Tor, пропонує технічно точне визначення: Dark Web - це наложена мережа, що працює поверх публічної інфраструктури Інтернету, але використовує спеціальні протоколи та порти, забезпечуючи анонімність та захист від аналізу трафіку [13].

1.2. Огляд джерел інформації для моніторингу

Основним джерелом інформації для моніторингу ресурсів Dark Web'у є спеціалізовані форуми та дискусійні платформи. Гарсія Херард зазначає, що: "Такі платформи, як Exploit.in, DarkWebForum і HackForums, часто слугують форумом для обміну інформацією про нові вразливості, методи атак і незаконну діяльність" [14]. Ретельний аналіз діяльності та дискусій на таких форумах може надати цінну інформацію, яка допоможе запобігти кіберзлочинам.

Наприклад, форум Exploit.in є одним з найбільших і найавторитетніших майданчиків у Dark Web'і, який спеціалізується на обговоренні вразливостей, методів експлуатації та інструментів для кібератак. На цьому форумі постійно публікуються повідомлення про нововиявлені вразливості в популярному програмному забезпеченні, детальні інструкції щодо їх використання, а також обговорюються нові тенденції та методи у сфері кіберзлочинності [15]. На Exploit.in постійно публікуються повідомлення про щойно виявлені вразливості в популярному програмному забезпеченні чи веб-сайтах (включно з Onion), а також докладні описи того, як їх використати. Учасники форуму активно діляться своїми знаннями про нові методи, інструменти і тактики, які можуть бути використані в кібератаках. Це включає в себе обговорення характеристик різних типів шкідливих програм, методів соціальної інженерії, злом акаунтів і схем крадіжки даних.

Також Exploit.in активно обговорюються нові тенденції в кіберзлочинності. Учасники форуму діляться досвідом використання нових інструментів і методик, обмінюються думками про розробку більш витончених методів атак і обговорюють найбільш ефективні способи уникнення виявлення правоохоронними органами.

DarkWebForum є платформою, де кіберзлочинці обмінюються інформацією про незаконні товари та послуги, доступні на темних веб-ринках. Тут можна знайти обговорення нових способів відмивання грошей, методів крадіжки даних, а також обміну досвідом щодо проведення успішних атак [16]. DarkWebForum - одна з найвідоміших платформ у Dark Web'і, де діє активна спільнота кіберзлочинців. На форумі поширюється інформація про нелегальні товари та послуги, доступні на ринку «темного інтернету».

Тут зустрічаються дискусії про нові методи відмивання грошей. Учасники діляться досвідом щодо найбільш ефективних способів викрадення та продажу вкрадених персональних даних, платіжних реквізитів та рахунків.

Крім того, на DarkWebForum активно обговорюються нові методи та інструменти для проведення кібератак. Кіберзлочинці діляться знаннями про вразливості програмного забезпечення, алгоритми обходу систем безпеки та

тактики соціальної інженерії для успішних атак на свої цілі.

Форум HackForums, хоча й не являється суто Dark Web-платформою, також становить інтерес для моніторингу, оскільки тут активно обговорюються новітні методи злому, написання шкідливого коду та інші техніки, які можуть бути використані у кіберзлочинній діяльності [17]. Форум HackForums, хоча й не належить до Dark Web'у, також становить інтерес для моніторингу та вивчення кіберзлочинності. Цей ресурс є популярним майданчиком для обговорення різноманітних аспектів хакерської діяльності. На HackForums активно обговорюються новітні методи злому різних систем та пристроїв. Учасники форуму діляться знаннями про уразливості у програмному забезпеченні, технології обходу систем безпеки та методи отримання несанкціонованого доступу до комп'ютерних мереж і баз даних.

Крім того, на цьому форумі можна знайти обмін інформацією щодо створення та поширення шкідливого програмного забезпечення. Тут публікуються приклади коду, інструкції з розробки та використання різноманітних видів зловмисних програм, таких як віруси, троянські коні, шпигунське ПЗ тощо.

Ретельний аналіз дискусій, повідомлень та активності користувачів на таких спеціалізованих форумах дозволяє виявляти ранні ознаки підготовки до кіберзлочинів, такі як:

- Обговорення нових вразливостей у програмному забезпеченні та методів їх експлуатації.
- Публікація інструкцій щодо створення та використання зловмисного програмного забезпечення.
- Обмін інформацією про незаконні товари та послуги, доступні на темних веб-ринках.
- Обговорення успішних кібератак та обмін досвідом їх проведення.
- Запити на пошук співучасників для організації кіберзлочинів.

Крім спеціалізованих форумів, важливим джерелом інформації для моніторингу темного веб-простору є самі темні веб-сайти та ринки, які пропонують широкий спектр нелегальних товарів і послуг. Відповідно до

дослідження Джонсона та Лі, моніторинг таких ресурсів може надати цінні розвідувальні дані, які можуть бути використані для запобігання кіберзлочинам [18].

Темні веб-ринки, такі як Silk Road, AlphaBay чи Hydra, являють собою онлайн-платформи в Dark Web'і, де продавці та покупці можуть анонімно здійснювати незаконні операції. Асортимент таких ринків включає наркотики, зброю, вкрадені дані, шкідливе програмне забезпечення та інші нелегальні товари. Моніторинг активності та пропозицій на цих майданчиках може виявити нові тенденції у сфері кіберзлочинності, такі як поява нових типів шкідливих програм, методів крадіжки даних або схем відмивання грошей [19] Детальніша характеристика веб-ринків наведена в таблиці 1.1.

Таблиця 1.1 – Порівняння Dark Web форумів

Показник	Silk Road	Hydra	AlphaBay
Час відкриття	2011 рік	2015 рік	2014 рік
Час закриття	2013 рік	2022 рік	2017 рік
Переваги (+)	+Першопроходець у сфері темних веб-ринків + Відносна простота використання та навігації	+ Широкий асортимент нелегальних товарів та послуг + Високий рівень анонімності	+ Велика різноманітність нелегальних товарів та послуг + Хороша репутація та довіра серед користувачів
Недоліки (-)	-Відносно невеликий асортимент товарів - Меншою мірою анонімний, ніж наступні ринки	- Високий рівень конкуренції між продавцями - Ризик шахрайства та крадіжки коштів	- Вразливість до правоохоронних операцій - Складніша у використанні, ніж інші ринки

Для безпечного доступу до темних веб-ресурсів дослідники зазвичай використовують анонімні мережі, такі як Tor або I2P, які приховують їхню ідентичність та місцезнаходження. Це дозволяє уникнути виявлення та

забезпечити конфіденційність під час моніторингу [20]. Однак особливої уваги заслуговують спеціалізовані пошукові системи, розроблені для навігації у мережі Dark Web'у. Алгоритм роботи представлений на рисунку 1.1.

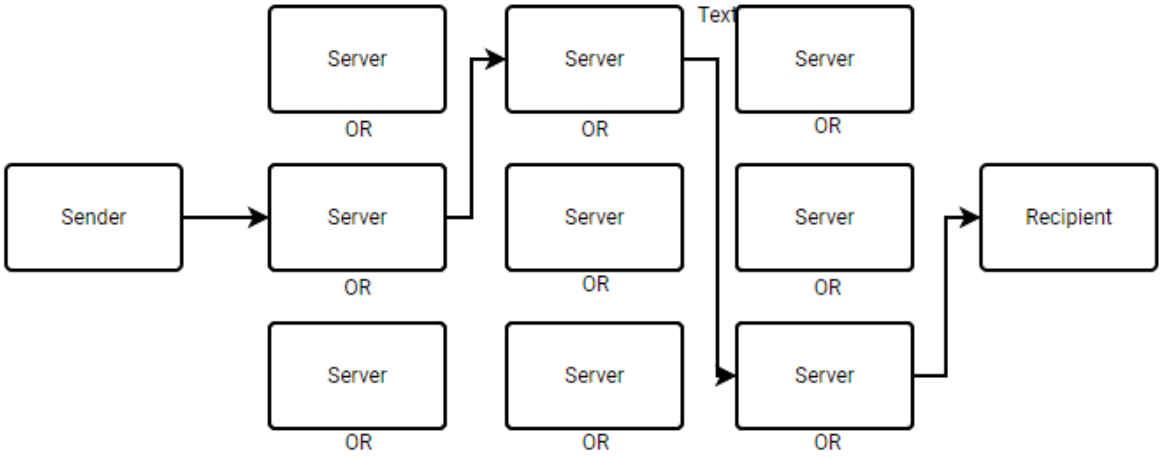


Рисунок 1.1 – Спрощена схема роботи Tor

Мережа Tor, відома також як "Dark Web", працює за принципом багаторівневого шифрування та випадкового вибору маршруту через свої вузли (реле). Коли користувач надсилає запит через Tor-браузер, його клієнт обирає довільний шлях з трьох випадкових реле у мережі Tor. Кожне реле знає лише адресу попереднього та наступного реле, але не знає початкового відправника чи кінцевого отримувача.

На кожному реле інформація додатково шифрується, а потім передається наступному реле. Таким чином, жодне з реле не має повної інформації про весь маршрут, лише частину. Коли запит досягає останнього, "вихідного" реле, воно розшифровує повідомлення та передає його в загальну мережу Інтернет. Отриманий ресурс потім прямує зворотнім шляхом через мережу Tor до початкового користувача.

Відповідно, через пошукові системи можна також знаходити інформацію, пробуючи відвідувати onion веб-сайти. Одна з ключових ролей у забезпеченні доступу до "прихованих" сайтів належить пошуковим системам, таким як Torch та DuckDuckGo. Ці платформи використовують складні алгоритми для індексації

та категоризації величезної кількості анонімних ресурсів Dark Web’у, що значно полегшує пошук та навігацію в його прихованій частині. На відміну від традиційних пошукових систем, вони спеціально адаптовані для роботи з "onion" адресами, що дозволяє користувачам отримувати доступ до широкого спектру інформації, товарів і послуг, доступних лише в Dark Web’і.

Пошукові системи, адаптовані для роботи з Dark Webом, використовують спеціальні технології, засновані на протоколі Tor, для доступу та індексації "onion" адрес. На відміну від традиційних пошукових систем, вони здатні виявляти та обходити обмеження, пов'язані з анонімністю та захищеністю Dark Web’у. Ці алгоритми дозволяють їм глибоко сканувати та індексувати величезну кількість ресурсів, які недоступні для звичайних пошукових платформ.

Після індексації, пошукові системи застосовують складні алгоритми для категоризації та структурування величезного масиву даних, отриманих з Dark Web’у. Вони групують ресурси за тематичними категоріями, такими як інформація, товари, послуги, форуми тощо. Ця систематизація значно полегшує навігацію користувачів та пошук необхідної інформації.

Ранжування та відображення результатів. Крім індексації та категоризації, пошукові системи використовують алгоритми для ранжування та відображення результатів пошуку. Вони враховують такі фактори, як релевантність, популярність, достовірність і безпеку ресурсів, щоб надати користувачам найбільш актуальну та значущу інформацію. Це дозволяє ефективно фільтрувати та представляти результати, полегшуючи пошук та взаємодію з ресурсами Dark Web’у.

Спеціалізовані каталоги, такі як The Hidden Wiki [21], також застосовують алгоритми для структурування та категоризації ресурсів Dark Weby. Вони збирають та систематизують величезну кількість посилань на "onion" сайти, групуючи їх за тематичними розділами. Ця організація ресурсів значно спрощує навігацію для користувачів, дозволяючи їм легко знаходити необхідну інформацію, товари або послуги.

Подібним чином, I2P (Invisible Internet Project) є анонімною децентралізованою мережею, що використовує власну маршрутизацію трафіку.

I2P-клієнт надає доступ до так званих "I2P-сайтів", які є аналогом звичайних веб-сайтів, але функціонують виключно в межах I2P-мережі. Завдяки цьому I2P також забезпечує високий рівень анонімності для дослідників темного веб-простору [22].

Freenet – ще один інструмент, що дозволяє безпечно та анонімно отримувати доступ до контенту в темному веб-просторі. На відміну від Tor та I2P, Freenet являє собою децентралізовану файлообмінну мережу, де інформація зберігається та передається анонімно між вузлами. Це робить Freenet більш стійким до цензури та моніторингу, але водночас і складнішим у використанні [23]. Як підкреслює Кларк, вміння ефективно використовувати ці інструменти є критично важливим для отримання релевантної інформації з темного веб-простору [24].

1.3. Кіберзлочинність у Dark Web, її види та особливості

Одним із поширених видів кіберзлочинності в Dark Web’і є торгівля шкідливими програмними засобами, такими як шкідливе ПЗ, троянські коні, віруси, а також набори для злому та злому систем. Ці інструменти часто пропонуються на спеціалізованих форумах і маркетплейсах, де зловмисники можуть придбати їх за криптовалюту, забезпечуючи при цьому анонімність транзакцій.

Ці інструменти включають, зокрема, троянські коні, віруси, черв'яки, програми-вимагачі, бекдори, шпигунське ПЗ, набори для взлому, скрипти та скрипти для автоматизації атак, програми для перебору паролів, інструменти для мережевого сканування та тестування на проникнення, а також програми для крадіжки паролів, облікових даних та особистої інформації.

Більш специфічними типами кіберзлочинності є торгівля нелегальними товарами та послугами, зокрема наркотиками, зброєю, підробленими документами, вкраденими даними, послугами кіберзлочинців чи навіть дитячою порнографією.

Щодо інформації, яку продають або дарують в Dark Web’і. Наприклад, у

травні та на початку червня 2022 року у Dark Web'і було викладено рекордну кількість баз даних російських компаній — понад 50. Для порівняння у квітні було 32, а у березні всього — 16. Серед жертв, чиї бази були викладені у публічний доступ, виявилися компанії, що належать до інтернет-сервісам доставки, медицині, телекому, інтернет-ритейлу, онлайн-освіті, будівництву та ін. Більшість баз даних популярних компаній були викладені в публічний доступ для безкоштовного скачування - зловмисники прагнуть не заробити, а завдати максимальних збитків компаніям та їх клієнтам.

За даними експертів Group-IB Threat Intelligence, найбільші 19 витоків у травні та на початку червня охопили 616,6 мільйона рядків. Майже всі бази даних містять імена клієнтів, номери телефонів, адреси, дати народження, деякі хеш-паролі, паспортні дані, деталі замовлень або результати медичних аналізів. Більшість баз даних були оновлені навесні цього року. На рисунку 1.2. і 1.3. зображено приклад такої бази даних, взятої з Dark Weby під час безплатної роздачі.

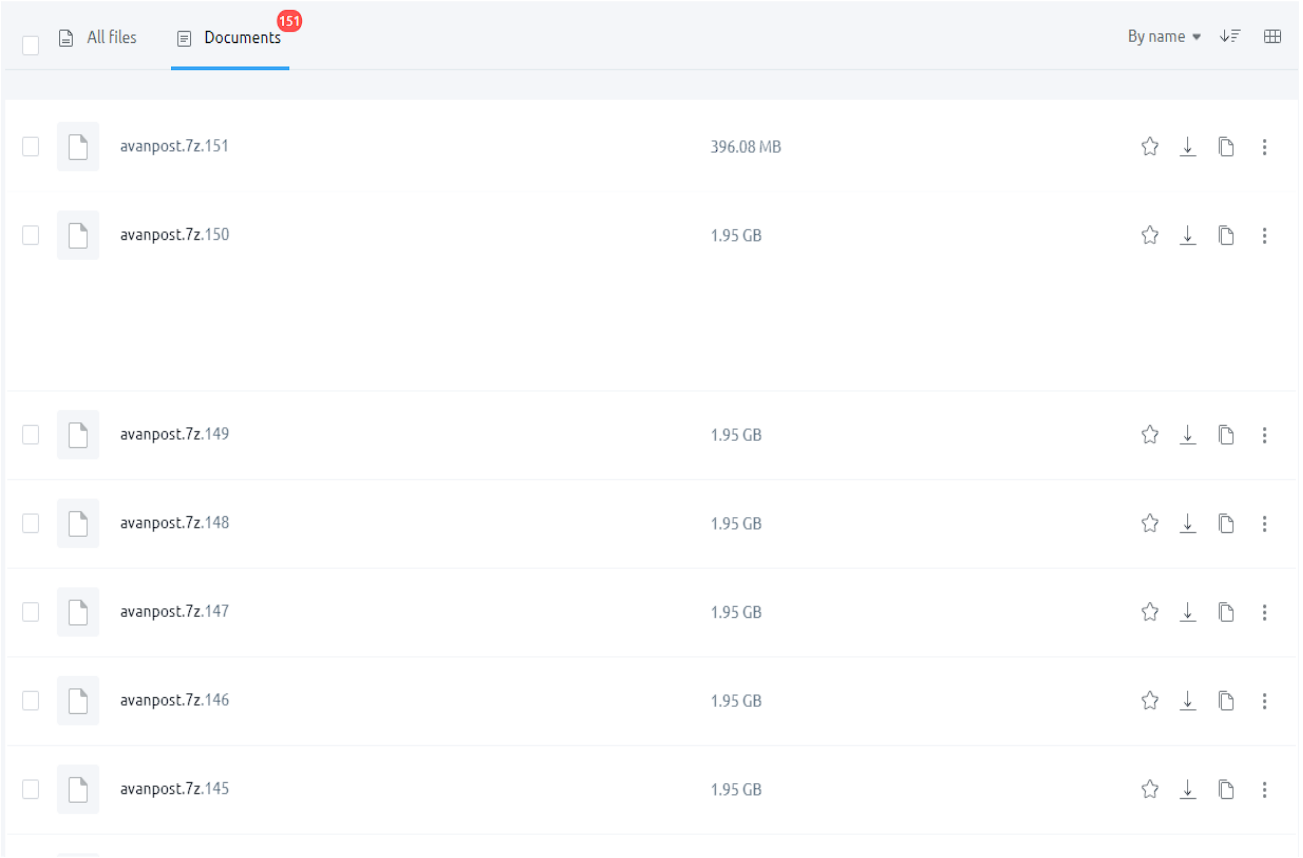
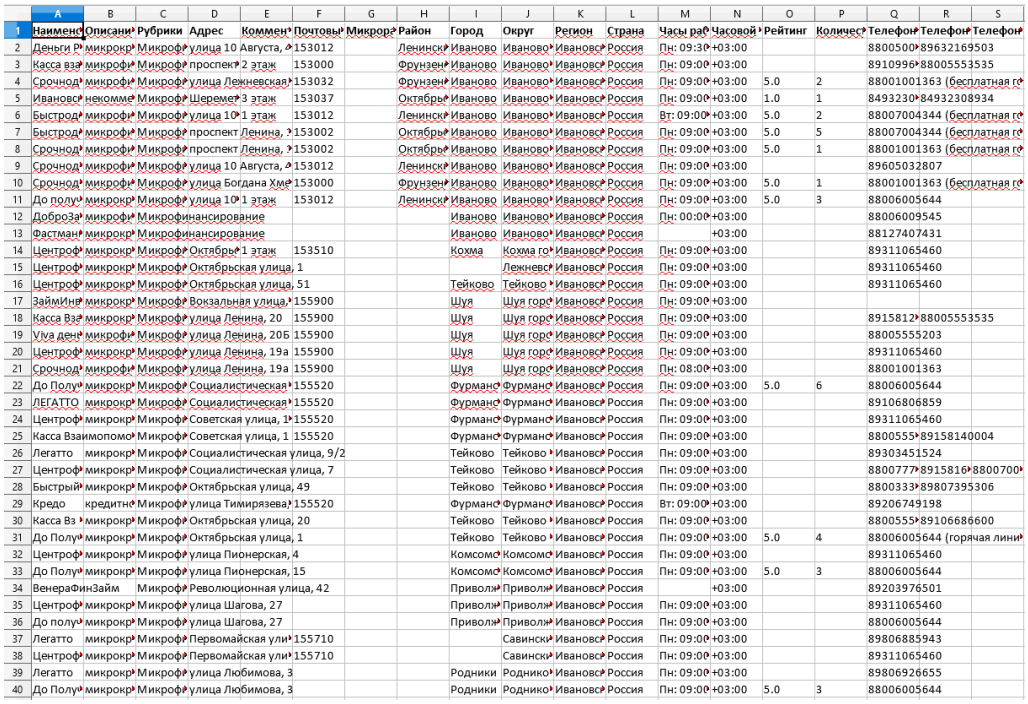


Рисунок 1.2 – Архів з базами даних під час безплатної роздачі

Відповідно, на рисунку 1.3. зображено вміст таких баз даних.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Наименов	Описани	Рубрики	Адрес	Коммен	Почтовы	Микрора	Район	Город	Округ	Регион	Страна	Часы ра	Часовой	Рейтинг	Количес	Телефон	Телефон	Телефон
2	Деньги Р	микрокр	Микрофи	улица 10 Августа,		153012		Ленинск	Иваново	Иваново	Ивановс	Россия	Пн: 09:30	+03:00			8800500	89632169503	
3	Касса вза	микрофи	Микрофи	проспект	2 этаж	153000		Фрунзен	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00			8910996	88005553535	
4	Срочнод	микрофи	Микрофи	улица Лежневская		153032		Фрунзен	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00	5.0	2	88001001363	(бесплатная го	
5	Ивановск	некомме	Микрофи	Шеремет	3 этаж	153037		Октябрь	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00	1.0	1	8493230	84932308934	
6	Быстрод	микрофи	Микрофи	улица 10	1 этаж	153012		Ленинск	Иваново	Иваново	Ивановс	Россия	Вт: 09:00	+03:00	5.0	2	88007004344	(бесплатная го	
7	Быстрод	микрофи	Микрофи	проспект Ленина,		153002		Октябрь	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00	5.0	5	88007004344	(бесплатная го	
8	Срочнод	микрофи	Микрофи	проспект Ленина,		153002		Октябрь	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00	5.0	1	88001001363	(бесплатная го	
9	Срочнод	микрофи	Микрофи	улица 10 Августа,		153012		Ленинск	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00			89605032807		
10	Срочнод	микрофи	Микрофи	улица Богдана Хме		153000		Фрунзен	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00	5.0	1	88001001363	(бесплатная го	
11	До полу	микрокр	Микрофи	улица 10	1 этаж	153012		Ленинск	Иваново	Иваново	Ивановс	Россия	Пн: 09:00	+03:00	5.0	3	88006005644		
12	ДоброЗа	микрофи	Микрофи	нансирование					Иваново	Иваново	Ивановс	Россия	Пн: 00:00	+03:00			88006009545		
13	Фастман	микрокр	Микрофи	нансирование					Иваново	Иваново	Ивановс	Россия		+03:00			88127407431		
14	Центроф	микрокр	Микрофи	Октябрь	1 этаж	153510			Кохма	Кохма го	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
15	Центроф	микрокр	Микрофи	Октябрьская улица,		1				Лежневс	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
16	Центроф	микрокр	Микрофи	Октябрьская улица,		51			Тейково	Тейково	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
17	ЗаймИнв	микрокр	Микрофи	Вокзальная улица,		155900			Шуя	Шуя горс	Ивановс	Россия	Пн: 09:00	+03:00					
18	Касса Вза	микрокр	Микрофи	улица Ленина,	20	155900			Шуя	Шуя горс	Ивановс	Россия	Пн: 09:00	+03:00			8915812	88005553535	
19	Viva дене	микрофи	Микрофи	улица Ленина,	20Б	155900			Шуя	Шуя горс	Ивановс	Россия	Пн: 09:00	+03:00			88005555203		
20	Центроф	микрокр	Микрофи	улица Ленина,	19а	155900			Шуя	Шуя горс	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
21	Срочнод	микрофи	Микрофи	улица Ленина,	19а	155900			Шуя	Шуя горс	Ивановс	Россия	Пн: 08:00	+03:00			88001001363		
22	До Полу	микрокр	Микрофи	Социалистическая		155520			Фурманс	Фурманс	Ивановс	Россия	Пн: 09:00	+03:00	5.0	6	88006005644		
23	ЛЕГАТТО	микрокр	Микрофи	Социалистическая		155520			Фурманс	Фурманс	Ивановс	Россия	Пн: 09:00	+03:00			89106806859		
24	Центроф	микрокр	Микрофи	Советская улица, 1		155520			Фурманс	Фурманс	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
25	Касса Вза	имопомо	Микрофи	Советская улица, 1		155520			Фурманс	Фурманс	Ивановс	Россия	Пн: 09:00	+03:00			8800555	89158140004	
26	Легатто	микрокр	Микрофи	Социалистическая улица, 9/2					Тейково	Тейково	Ивановс	Россия	Пн: 09:00	+03:00			89303451524		
27	Центроф	микрокр	Микрофи	Социалистическая улица, 7					Тейково	Тейково	Ивановс	Россия	Пн: 09:00	+03:00			8800777	8915816	8800700
28	Быстрый	микрокр	Микрофи	Октябрьская улица, 49					Тейково	Тейково	Ивановс	Россия	Пн: 09:00	+03:00			8800333	89807395306	
29	Кредо	кредитно	Микрофи	улица Тимирязева,		155520			Фурманс	Фурманс	Ивановс	Россия	Вт: 09:00	+03:00			89206749198		
30	Касса Вз	микрокр	Микрофи	Октябрьская улица,		20			Тейково	Тейково	Ивановс	Россия	Пн: 09:00	+03:00			8800555	89106686600	
31	До Полу	микрокр	Микрофи	Октябрьская улица,		1			Тейково	Тейково	Ивановс	Россия	Пн: 09:00	+03:00	5.0	4	88006005644	(горячая лини	
32	Центроф	микрокр	Микрофи	улица Пионерская,		4			Комсомс	Комсомс	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
33	До Полу	микрокр	Микрофи	улица Пионерская,		15			Комсомс	Комсомс	Ивановс	Россия	Пн: 09:00	+03:00	5.0	3	88006005644		
34	ВенераФин	Займ	Микрофи	Революционная улица,		42			Приволж	Приволж	Ивановс	Россия		+03:00			89203976501		
35	Центроф	микрокр	Микрофи	улица Шагова,		27			Приволж	Приволж	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
36	До полу	микрокр	Микрофи	улица Шагова,		27			Приволж	Приволж	Ивановс	Россия	Пн: 09:00	+03:00			88006005644		
37	Легатто	микрокр	Микрофи	Первомайская ули		155710				Савинск	Ивановс	Россия	Пн: 09:00	+03:00			89806885943		
38	Центроф	микрокр	Микрофи	Первомайская ули		155710				Савинск	Ивановс	Россия	Пн: 09:00	+03:00			89311065460		
39	Легатто	микрокр	Микрофи	улица Любимова,		3			Родники	Роднико	Ивановс	Россия	Пн: 09:00	+03:00			89806926655		
40	До Полу	микрокр	Микрофи	улица Любимова,		3			Родники	Роднико	Ивановс	Россия	Пн: 09:00	+03:00	5.0	3	88006005644		

Рисунок 1.3 - Одна з баз даних

Мотив зловмисника - не заробити гроші, а завдати якомога більшої шкоди компанії [25]. На думку експертів департаменту Threat Intelligence Group-IB, проблема такої величезної кількості інцидентів, крім очевидних витоків - дії інсайдерів і збільшення кількості кібератак з 24 лютого.

Наприкінці квітня в рамках глобального дослідження цифрових активів Group-IB було виявлено близько 400 000 загальнодоступних баз даних, що зберігаються у відкритому доступі. Майже 7 500 з них були на російських серверах [26].

Більше половини інцидентів, розслідуваних лабораторією цифрової криміналістики Group-IB у 2021 році, сталися внаслідок експлуатації вразливостей периметра. Нове рішення Group-IB для управління поверхнею атаки (Attack Surface Management) в режимі реального часу дозволяє робити швидкий аналіз ризиків для компанії: чи викладені дані в Dark Web'і, чи є відкриті порти, чи є вразливості, які можуть бути експлуатовані та інші.

Стосовно інших перспективних напрямів, дослідження проводила Privacy Affairs. Дослідження показало, що обсяг витоку даних, що продаються в Dark Web'і, значно збільшився порівняно з 2021 роком. Особливою популярністю

користуються підроблені посвідчення особи та номери кредитних карток. Зросла не тільки кількість, а й різноманітність товарів для покупки, наприклад, зламані криптовалютні рахунки та веб-сервіси, такі як облікові записи Uber. На рисунку 1.4. зображено виписку популярних документіві для продажу в Dark Web'і за 2021 рік.

Rating	Listing Title	User Comment	Date
5/5	(Premium) North Carolina Fake ID/Drivers License w/ Tracking (NC)	None left	2021-02-02
Rating	**Listing Title**	**User Comment**	**Date**
5/5	(Quality) New Jersey Fake ID/Drivers License w/ Tracking (NJ)	great id. will be back	2021-02-01
Rating	**Listing Title**	**User Comment**	**Date**
5/5	(Quality) Missouri Fake ID/Drivers License w/ Tracking (MO)	ids are amazing	2021-02-01
Rating	**Listing Title**	**User Comment**	**Date**
5/5	(Quality) Pennsylvania Fake ID/Drivers License w/ Tracking (PA)	This is my 7th purchase with this vendor and i received the id everytime quickly and the work is actually better then qualityfakeids And the jackass that said this vendor sent a book thats impossle just another LIEING cheapskate trying to get this vendors work for free BUY WITH CONFIDENCE, BEST FAKE REAL IDS ON HERE ! Infact my first ever buy was not on here so this vendor can be trusted off here	2021-01-30
Rating	**Listing Title**	**User Comment**	**Date**
5/5	(Quality) Indiana Fake ID/Drivers License w/ Tracking (IN)	a1	2021-01-30
Rating	**Listing Title**	**User Comment**	**Date**
5/5	(Premium) Texas Fake ID/Drivers License w/ Tracking (TX)	goated	2021-01-29

Рисунок 1.4 – Виписка популярних документів в Dark Web`i

Незважаючи на зростання пропозиції, ціни на клоновані кредитні картки та пов'язані з ними дані зросли в ціні порівняно з минулим роком. Швидше за все це зумовлено кількома факторами: збільшення ризику отримання інформації, збільшення вигоди покупців від використання інформації, підвищення якості/точності даних карток або результат дії після 24 лютого 2022 року, після чого виросла необхідність в певних махінаціях зі сторони держави.

Детальніше про ціни можна подивитись в таблиці 1.2., сформованій на основі дослідження Privacy Affairs.

Таблиця 1.2 – Продукти Dark Web'у та їх середня ціна

Категорія	Продукт	Середня ціна Dark Weby (USD)
Дані кредитної картки	Реквізити кредитної картки, залишок на рахунку до 5 000	110 дол
	Реквізити кредитної картки, залишок на рахунку до 1 000	70 дол
	Вкрадені логіни онлайн-банкінгу, мінімум 2 000 на рахунку	60 дол.
Послуги з обробки платежів	Вхід в банківський рахунок ING (підтверджений обліковий запис)	4 255 дол.
	Бізнес-аккаунт HSBC у Великій Британії	4 200 дол
	Вхід в інтернет-банкінг Швейцарії	2 200 дол.
	Вхід в онлайн-банкінг Barclays	2,100 дол.
	Особистий банківський рахунок Santander	1 800 дол.
Криптовалютні рахунки	Верифікований обліковий запис N26 (Німеччина)	2 650 дол.
	Верифікований та зламаний акаунт Wirex	2,300 дол.

Ще одним ринком, який є поширений в Dark Web'і - наркотики та зброя. Так з початком війни, в Dark Web'і часто почали з'являтись магазини, які будцім продають зброю зі Заходу, яку надають Україні. Так у Dark Web'і один із "продавців", який часто згадується в ЗМІ, називається Weapons Ukraine. На сторінці магазину є згадка про 32 нібито проведені ним угоди з продажу американських карабінів, пістолетів та іншої вогнепальної зброї, яку нібито доставляють анонімним покупцям по всій території України. На рисунку 1.5. зображено ніби-то сайт з продажем зброї.

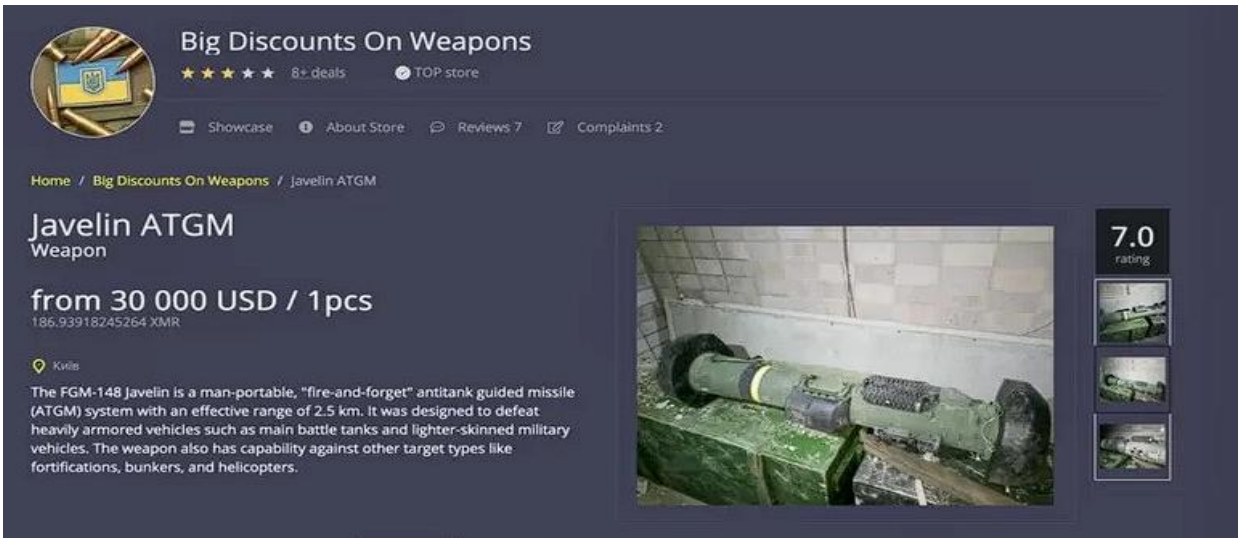


Рисунок 1.5 – Продаж зброї у Dark Web

Хоча варто зазначити, що такий продаж - звичайна махінація зі сторони недобросовісних користувачів.

Також популярним товаром серед Dark Web'у є дитяча порнографія. Одним з найпопулярніших був Lolita City - веб-сайт, що знаходився в зоні .onion анонімної мережі Tor. Сайт був онлайн-галереєю дитячої порнографії з фотографіями моделей жіночої та чоловічої статі від дитячого до 17-річного віку (18 років є мінімальним легальним віком моделей для зйомок у порнографії в більшості країн світу) [27].

Окрім товарів, в Dark Web'і також продаються і послуги. Це може бути вбивство, наприклад. За словами правоохоронців, вартість послуги варіюється в залежності від способу та обставин вбивства, а також кого саме потрібно усунути. Для зловмисників є різниця між підприємцем та колишнім кримінальним авторитетом. Найяскравіший випадок такого - замовлення вбивства в Китаї, де один і той самий контракт забрали 5 людей зі зменшенням вартості.

Суд Китаю засудив шістьох обвинувачених в організації замаху місцевого підприємця до тюремних термінів. TJ називає цю історію «сагою про аутсорсинг».

Все почалося в 2013 році, коли власник ріелторської компанії Тан Юхуей найняв кілера, щоб позбутися конкурента на прізвище Вей. Бізнесмен заплатив

кілеру два мільйони юанів (приблизно 18 мільйонів рублів) готівкою, але той узяв собі половину суми, а другу віддав іншому вбивці.

Наступний найманець зробив так само — четвертому кілеру за вбивство запропонували 100 тисяч юанів. [28]. Тобто, ціни в Dark Web'і на вбивство можуть варіюватись.

1.4. Правові аспекти використання джерел для моніторингу Dark Web'у

В Україні не існує як такого закону про регулювання моніторингу, проте питання використання джерел для моніторингу та дослідження інтернету, що і стосується Dark Web'у, існує. Воно регулюється низкою законодавчих актів, що охоплюють сфери кіберзлочинності, захисту персональних даних та інформаційної безпеки. Ці нормативно-правові документи встановлюють важливі рамки для здійснення такої діяльності з урахуванням ключових правових принципів.

Закон України "Про основні засади забезпечення кібербезпеки України" [29] є ключовим нормативно-правовим актом, що визначає правові основи протидії кіберзлочинності в Україні. Цей законодавчий документ детально регламентує повноваження та функції уповноважених державних органів у сфері виявлення, запобігання та розслідування кіберінцидентів, у тому числі тих, що можуть бути пов'язані з діяльністю у Dark Web'і.

Зокрема, закон наділяє компетентні органи повноваженнями щодо моніторингу кіберпростору, збору та аналізу інформації про кіберзагрози, а також вжиття заходів реагування на кіберінциденти [30]. Ці повноваження є критично важливими для протидії кіберзлочинності, що проявляється у Dark Web'і, оскільки вони дозволяють уповноваженим органам здійснювати цілеспрямований пошук, виявлення та документування протиправної діяльності в цій приховані частині Інтернету.

Водночас, закон підкреслює необхідність дотримання балансу між забезпеченням кібербезпеки та захистом основоположних прав і свобод людини, зокрема права на приватність [31]. Ця вимога є важливою, оскільки моніторинг

Dark Weby може потенційно зачіпати конфіденційність користувачів, які, хоча й здійснюють протиправну діяльність, також можуть використовувати Dark Web для законних цілей.

Для досягнення цього балансу, закон передбачає механізми співпраці між державними органами, операторами критичної інфраструктури, провайдерами електронних комунікацій та іншими суб'єктами сфери кібербезпеки [32]. Така взаємодія дозволяє консолідувати зусилля, обмінюватися інформацією та вживати скоординованих заходів для ефективного виявлення, запобігання та реагування на кіберзагрози, у тому числі ті, що пов'язані з діяльністю у Dark Web'і. Закон встановлює вимоги до забезпечення кібербезпеки об'єктів критичної інфраструктури [33].

Крім того, Закон України "Про захист персональних даних" встановлює вимоги щодо правомірності обробки персональних даних, у тому числі отриманих з ресурсів Dark Web'у. Цей нормативно-правовий акт покладає на суб'єктів, що здійснюють моніторинг Dark Web'у, обов'язок забезпечувати конфіденційність та безпеку персональних даних відповідно до визначених законодавством гарантій.

Цей закон встановлює вимоги щодо правомірності обробки персональних даних, у тому числі тих, що можуть бути отримані з ресурсів Dark Web'у під час проведення моніторингу та розслідувань [34]. Він покладає на суб'єктів, які здійснюють таку діяльність, чіткі зобов'язання щодо забезпечення конфіденційності та безпеки персональних даних відповідно до визначених законодавством гарантій.

Закон вимагає від операторів персональних даних, до яких можуть належати органи, що здійснюють моніторинг Dark Weby, дотримуватися принципів законності, справедливості та прозорості при обробці персональних даних [35]. Вони також зобов'язані забезпечити захист персональних даних від несанкціонованого доступу, знищення, модифікації, блокування, копіювання, поширення та інших неправомірних дій [36].

Важливим аспектом правового регулювання є також Закон України "Про інформацію", який регулює порядок доступу до інформації, її збирання,

використання та поширення. Ці положення мають важливе значення при здійсненні моніторингу Dark Web’у, оскільки визначають умови легального отримання, обробки та розкриття відомостей, що можуть бути отримані з його ресурсів.

Цей закон визначає правові режими доступу до інформації, а також порядок її збирання, використання та поширення [37]. Ці положення мають суттєве значення при здійсненні моніторингу Dark Web’у, оскільки вони встановлюють умови легального отримання, обробки та розкриття відомостей, що можуть бути отримані з ресурсів Dark Web’у.

Зокрема, закон встановлює принципи доступу до інформації, серед яких - презумпція відкритості інформації, обмеження доступу виключно на підставах, визначених законом, та забезпечення права кожного на доступ до інформації [38]. Ці принципи є важливими при проведенні моніторингу Dark Web’у, адже вони визначають межі правомірного збору та використання даних з його ресурсів.

Крім того, Кримінальний кодекс України містить статті, що передбачають відповідальність за незаконні дії, пов'язані з використанням комп'ютерних систем, мереж та електронних носіїв інформації. Кримінальний кодекс України [39] містить низку статей, які мають важливе значення при визначенні правомірності методів моніторингу Dark Web’у, особливо в контексті збору доказової бази для кримінальних проваджень.

Наприклад, Розділ XVI Кримінального кодексу "Кримінальні правопорушення у сфері використання електронно-обчислювальних машин (комп'ютерів), систем та комп'ютерних мереж і мереж електрозв'язку" встановлює відповідальність за низку злочинів, пов'язаних із незаконним втручанням у роботу комп'ютерних систем, незаконним доступом до інформації, створенням і розповсюдженням шкідливих програмних засобів тощо [40]. Ці норми є важливими при оцінці законності методів, використовуваних правоохоронними органами та уповноваженими суб'єктами для моніторингу Dark Web’у.

Так, стаття 361 Кримінального кодексу "Несанкціоноване втручання в

роботу електронно-обчислювальних машин (комп'ютерів), автоматизованих систем, комп'ютерних мереж чи мереж електрозв'язку" встановлює відповідальність за незаконне втручання в роботу комп'ютерних систем, що може мати місце при проведенні моніторингу Dark Web'у [41].

Ця стаття може бути зв'язана з людиною, якщо в подальшому згідно втручання було застосовано машину для використання її і розповсюдження визначеного матеріалу згідно статті 301 чи інших статей ККУ.

Стаття 361-1 Кримінального кодексу "Створення з метою використання, розповсюдження або збуту шкідливих програмних чи технічних засобів, а також їх розповсюдження або збут" передбачає відповідальність за дії, пов'язані із створенням і поширенням шкідливого програмного забезпечення [42].

Ця норма може бути застосована до осіб, які використовують такі інструменти для протиправної діяльності в Dark Web'і, що підлягає виявленню та документуванню правоохоронними органами.

В таблиці 1.3. сформовано основні відповідності між законами України та їх відповідністю з Dark Web'ом, а також підв'язано застосування статті кримінального кодексу України до можливих правопорушень в мережі Dark Web'у з тими, які застосовуються кіберполіцією і для звичайних електронно-обчислювальних машин.

Таблиця 1.3 – Статті ККУ зі застосуванням до Dark Web'у

Стаття ККУ	Назва статті	Застосування до Dark Web
Стаття 182	Порушення недоторканності приватного життя	Може застосовуватися до незаконного доступу, збору, використання або поширення персональних даних користувачів Dark Web
Стаття 191	Привласнення, розтрата майна або заволодіння ним шляхом зловживання службовим становищем	Може застосовуватися до незаконного збагачення за рахунок протиправної діяльності в Dark Web

Продовження таблиці 1.3

Стаття 361	Несанкціоноване втручання в роботу електронно-обчислювальних машин (комп'ютерів), автоматизованих систем, комп'ютерних мереж чи мереж електрозв'язку	Безпосередньо застосовується до незаконного втручання в роботу комп'ютерних систем, для застосування його для подальшого хостингу або продажу в Dark Web.
Стаття 361-1	Створення з метою використання, розповсюдження або збуту шкідливих програмних чи технічних засобів, а також їх розповсюдження або збут	Застосовується до виготовлення та поширення шкідливого програмного забезпечення, що використовується в Dark Web
Стаття 301-1	Ввезення, виготовлення, збут і розповсюдження порнографічних предметів	Застосовується до виготовлення, збуту та розповсюдження дитячої порнографії, що може мати місце в Dark Web

Якщо розглядати інші країни, то існують і виключення. Деякі країни, такі як Ісландія та Німеччина, займають відносно ліберальну позицію щодо використання Dark Web. В Ісландії, наприклад, використання Tor та інших анонімних мереж не заборонено законом, доки воно не пов'язане з незаконною діяльністю [44]. Німеччина також не забороняє використання Dark Web, але бореться з конкретними злочинами, що вчиняються через нього [45].

Інші країни, як-от Росія та Китай, мають більш обмежувальне законодавство щодо Dark Web. Росія, наприклад, заборонила використання Tor та інших анонімних мереж, окрім як для державних установ [46]. Китай також посилив контроль над доступом до Dark Web, включаючи блокування відповідних браузерів та VPN-сервісів [47].

Незалежно від загального ставлення до Dark Web, більшість країн світу криміналізують певні види кіберзлочинності, пов'язані з діяльністю в Dark Web. Зокрема, широко розповсюджені норми, що встановлюють відповідальність за незаконний доступ до комп'ютерних систем, розповсюдження шкідливого ПЗ, торгівлю нелегальними товарами тощо [48].

На міжнародному рівні діє Конвенція Ради Європи про кіберзлочинність, яка є основним правовим документом, що регулює боротьбу з транскордонною кіберзлочинністю, у тому числі пов'язаною з Dark Web'ом. Країни-учасниці зобов'язуються гармонізувати національне законодавство та співпрацювати у сфері розслідування та переслідування таких злочинів [49].

На міжнародному рівні ключовим правовим документом, що регулює боротьбу з транскордонною кіберзлочинністю, у тому числі пов'язаною з діяльністю в Dark Web, є Конвенція Ради Європи про кіберзлочинність [50]. Ця конвенція, яку також називають Будапештською конвенцією, є першим міжнародним договором, спрямованим на гармонізацію національного законодавства та сприяння міжнародному співробітництву у сфері протидії кіберзлочинності.

Ключовим аспектом Конвенції є її екстериторіальний характер. Вона встановлює зобов'язання для країн-учасниць щодо криміналізації широкого спектру кіберзлочинів, незалежно від того, де вони були вчинені чи виявлені [51]. Таким чином, навіть злочини, пов'язані з використанням Dark Web'у, можуть підлягати переслідуванню відповідно до положень цієї Конвенції.

Конвенція визначає чотири основні категорії кіберзлочинів, на які поширюється її дія: злочини проти конфіденційності, цілісності та доступності комп'ютерних даних і систем; злочини, пов'язані з використанням комп'ютерів; злочини, пов'язані зі змістом; та злочини, пов'язані з порушенням авторських і суміжних прав [52]. Ці категорії охоплюють широкий спектр протиправних дій, які можуть мати місце в Dark Web, від незаконного доступу до комп'ютерних систем до розповсюдження шкідливого контенту.

Для забезпечення ефективності боротьби з транскордонною кіберзлочинністю, Конвенція також встановлює механізми міжнародного співробітництва. Вона зобов'язує країни-учасниці надавати взаємну правову допомогу у розслідуванні та переслідуванні кіберзлочинів, а також запроваджує процедури термінового збереження комп'ютерних даних, перехоплення трафіку та отримання доступу до збережених комп'ютерних даних [53]. Ці положення є критично важливими для ефективного реагування на злочини, пов'язані з

використанням анонімних мереж, таких як Dark Web. Конвенція також приділяє увагу захисту прав людини та основоположних свобод при застосуванні заходів боротьби з кіберзлочинністю. Вона наголошує на необхідності дотримання принципів пропорційності та процесуальних гарантій, а також забезпечення балансу між забезпеченням безпеки та захистом приватності [54].

Висновки до розділу 1.

1. Проведено теоретичний аналіз Dark Web та діпнету, розглянуто їх сутність, характеристики та основні компоненти, а також проаналізовано підходи інших експертів до визначення цих понять.

2. Визначено та проаналізовано основні джерела інформації, що потребують моніторингу для ефективного виявлення та запобігання кіберзлочинності у Dark Web.

3. Проведено детальний аналіз основних видів кіберзлочинності в Dark Web, розглянуто специфіку та різновиди послуг і товарів.

4. Проаналізовано правові аспекти України використання джерел для моніторингу Dark Web, також розглянуто закони інших держав.

2 МЕТОДИ ТА ЗАСОБИ МОНІТОРИНГУ ТЕМНИХ ВЕБ-РЕСУРСІВ

2.1. Технології анонімного доступу та безпеки в Dark Web

Для реалізації програмного продукту необхідно розпочати з постановки ідеї програми. Суть роботи полягає в тому, щоб розробити алгоритм програми, яка отримує доступ в Dark Web. Для початку потрібно зрозуміти, як саме працюють onion лінки та як виглядає вхід без допомоги автоматизації.

Для доступу до Dark Web необхідно почати з завантаження та налаштування спеціалізованого програмного забезпечення, що дозволяє обійти звичайні способи підключення до інтернету та забезпечує анонімність користувача. Однією з найпоширеніших платформ для доступу до Dark Web є мережа Tor (The Onion Router). Першим кроком є завантаження браузера Tor, який є модифікованою версією Firefox та дозволяє користувачам анонімно переглядати сайти з доменами ".onion". На рисунку 2.1. зображено сайт, з якого завантажено Tor браузер. Варто зазначити, що Tor існує для IOS, Mac, Android та Windows, проте протягом виконання роботи було використано саме екземпляр для Windows.

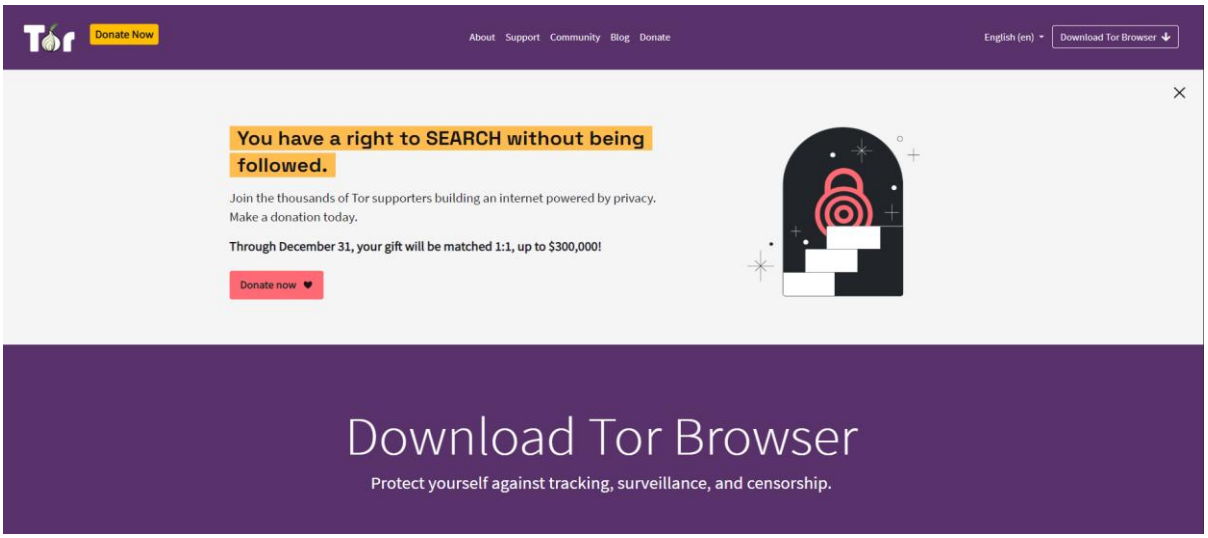


Рисунок 2.1 – Сторінка завантаження Tor

Після завантаження браузера Tor з офіційного сайту (https://www.torproject.org/), його потрібно встановити на комп'ютері. Під час інсталяції браузер не вимагає особливих налаштувань, але рекомендовано на

форумі зберігати його в окремій папці, щоб не можна було отримати доступ по стандартним лінкам у випадку, коли комп’ютер перевіряють на встроєні браузери. Після встановлення необхідно запустити програму та натиснути кнопку "Connect", що дозволить браузеру встановити з'єднання з мережею Tor. Це з'єднання відбувається через кілька проміжних серверів, які шифрують дані користувача, приховуючи його реальну IP-адресу та місцезнаходження. Цей процес може займати кілька секунд, залежно від швидкості інтернет-з'єднання. На рисунку 2.2. зображено алгоритм приховування IP адреси та з’єднання користувача з кінцевим сайтом.

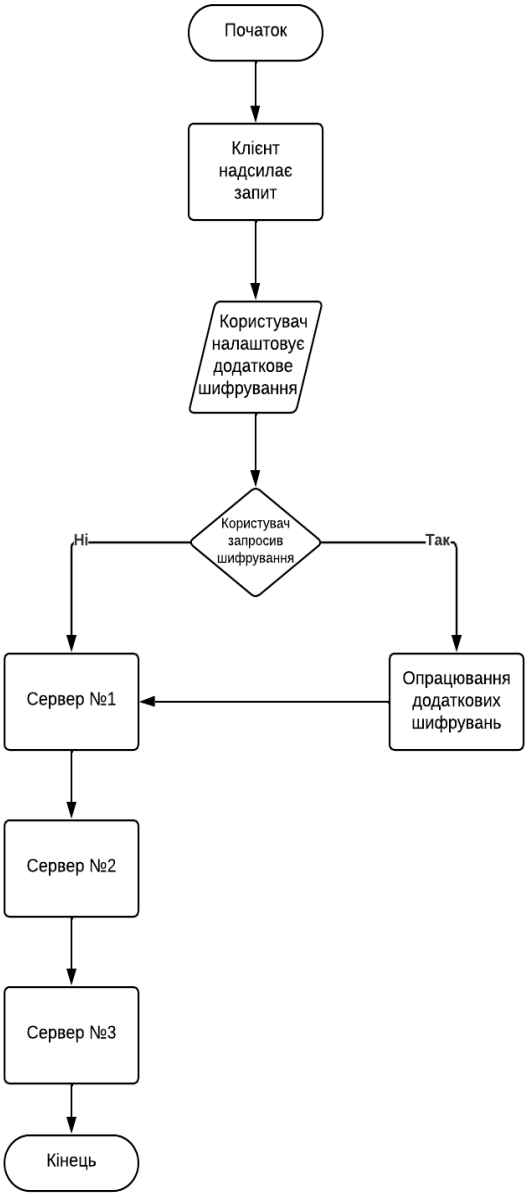


Рисунок 2.2 – Схема алгоритму з’єднання Tor браузера з кінцевим хостом

Tor автоматично надає 3 прослойки для захисту користувача в мережі, щоб важче було відслідкувати кінцевий шлях. Після цього будь-який користувач може виконувати пошукові запити навіть за звичайними пошуковими системами, типу bing, google.

У Dark Web можна знайти різноманітні ресурси, від форумів та блогів до торгових платформ і бібліотек. Однак, через специфіку мережі, користувачам слід бути готовими до нестабільної роботи сайтів. Часто ресурси можуть бути недоступними через технічні проблеми або блокування, а деякі з них можуть змінювати свої URL-адреси для уникнення відстеження. Також потрібно враховувати, що анонімність у Dark Web не є абсолютною. Використання плагінів, відкриття документів або завантаження файлів безпосередньо через браузер Tor можуть розкрити IP-адресу або іншу особисту інформацію, тому для забезпечення додаткової безпеки та конфіденційності, під час використання Dark Web варто дотримуватися кількох правил.

По-перше, не рекомендується вводити будь-яку особисту інформацію, таку як ім'я, електронну пошту чи паролі, особливо на незнайомих сайтах. Крім того, варто уникати завантаження файлів з неперевірених джерел, оскільки вони можуть містити шкідливе програмне забезпечення.

По-друге, щоб підвищити рівень безпеки, користувачам рекомендується використовувати VPN (Virtual Private Network) разом із браузером Tor, що дозволяє додатково шифрувати трафік і ще більше ускладнити відстеження діяльності в мережі. На рисунку 2.3. зображено алгоритм роботи VPN. VPN являє собою комплексну мережеву технологію, що забезпечує створення захищеного віртуального каналу між пристроєм користувача та цільовим сервером через публічну мережу інтернет. Основу роботи VPN складає багаторівнева система шифрування даних та тунелювання трафіку, що дозволяє приховати реальну IP-адресу користувача та захистити його мережеву активність від стороннього спостереження. При підключенні до VPN-сервера встановлюється захищене з'єднання з використанням протоколів шифрування, таких як IPSec, OpenVPN або WireGuard, які забезпечують конфіденційність та цілісність даних при їх передачі через публічні мережі.

У процесі роботи VPN-з'єднання відбувається автоматична інкапсуляція мережевих пакетів, при якій весь трафік користувача спочатку шифрується на пристрої, після чого упаковується в нові пакети з оновленими заголовками для маршрутизації через менеджер конфігурації. Такий підхід забезпечує подвійний захист - шифрування самих даних та приховування реальних мережевих адрес. VPN-сервер, отримуючи зашифровані пакети, розшифровує їх та направляє до кінцевого пункту призначення від свого імені, виступаючи посередником між користувачем та цільовими ресурсами. При цьому для зовнішніх спостерігачів весь трафік виглядає як такий, що надходить від VPN-сервера, а не від реального користувача.

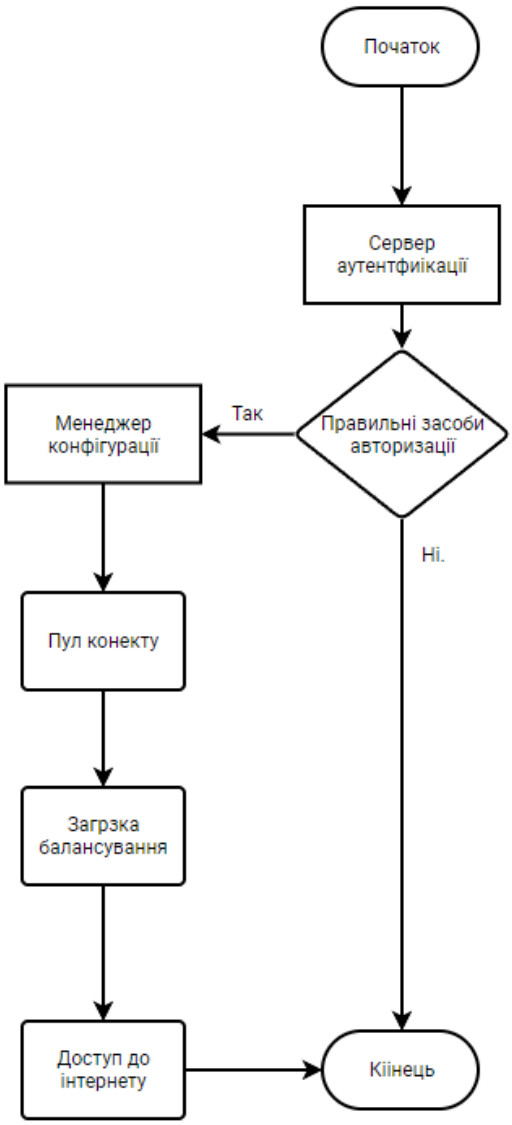


Рисунок 2.3 – Алгоритм роботи VPN

Модуль Connection Pool у VPN-системах забезпечує ефективне управління множинними клієнтськими підключеннями та оптимізацію використання серверних ресурсів. Пул підключень реалізує механізм кешування та повторного використання мережевих з'єднань, що дозволяє уникнути надмірних витрат на встановлення нових сесій для кожного запиту. При надходженні запиту від клієнта система спершу перевіряє наявність вільних з'єднань у пулі, і якщо такі існують - надає їх для обробки запиту, замість створення нового підключення. Це суттєво знижує затримки та навантаження на сервер, оскільки процес встановлення нового захищеного з'єднання вимагає значних обчислювальних ресурсів для генерації ключів шифрування та виконання процедур аутентифікації.

Connection Pool підтримує динамічне масштабування кількості активних підключень відповідно до поточного навантаження на систему. При збільшенні числа клієнтських запитів пул автоматично створює додаткові з'єднання до досягнення встановленого максимального ліміту. Коли навантаження знижується, надлишкові з'єднання поступово закриваються для вивільнення системних ресурсів. Такий адаптивний підхід забезпечує оптимальний баланс між доступністю сервісу та ефективністю використання ресурсів сервера. Модуль також відстежує стан кожного з'єднання в пулі, періодично перевіряючи їх працездатність та автоматично відновлюючи з'єднання, що були розірвані через мережеві проблеми.

Важливою функцією Connection Pool є реалізація механізмів балансування навантаження між наявними з'єднаннями. Система використовує різні алгоритми розподілу запитів, такі як Round Robin, Least Connections або Weight-based, для рівномірного розподілу навантаження між активними з'єднаннями. Це дозволяє максимально ефективно використовувати доступні ресурси та запобігати перевантаженню окремих каналів зв'язку. Додатково реалізується механізм пріоритизації запитів, який забезпечує першочергову обробку критично важливих операцій навіть при високому загальному навантаженні на систему. На рисунку 2.4. зображено що міститься всередині connection_pool на основі того, що вже було описано вище.

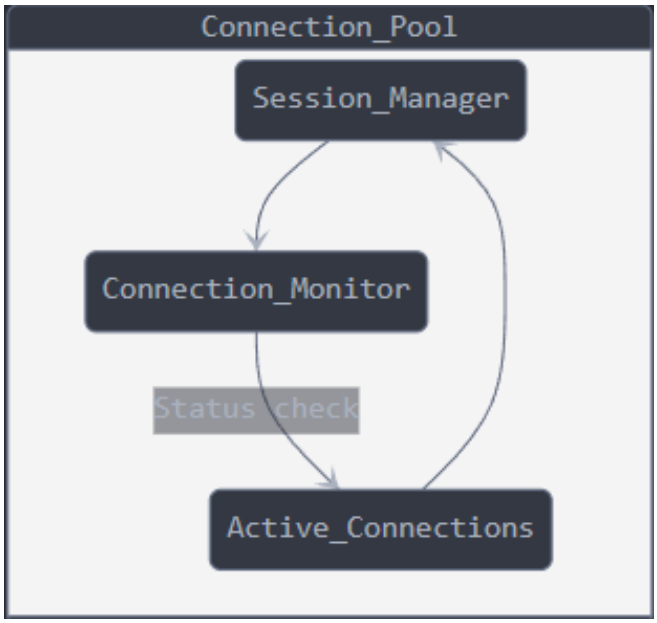


Рисунок 2.4 – Схема взаємодії модулів в Connection_pool

Саме цей модуль відправляє до шифрування даних та перевірки ключів. В деяких випадках використовують proxy_server, проте це гірше, аніж VPN, якщо використовувати Tor за призначенням. Проксі-сервер принципово відрізняється від vpn в залежності від механізму його роботи і рівня захисту призначених для користувача даних. Проксі-сервер працює на рівні окремого додатка або протоколу і діє як посередник між клієнтом і цільовим сервером для певного типу з'єднання. При використанні проксі-деки трафік маршрутизується тільки через проміжний сервер, і не створюються безпечні тунелі або складне шифрування даних. Проксі-сервер може працювати з різними протоколами, такими як HTTP, SOCKS та FTP, але кожен Тип проксі-сервера, як правило, обмежений підтримкою певного протоколу.

На відміну від VPN, проксі-сервери не забезпечують захист системного рівня і не інтегруються з мережевим стеком операційної системи. Це означає, що, хоча користувачам потрібно індивідуально налаштувати кожну програму для роботи через проксі-сервер, VPN автоматично захищає весь мережевий трафік на пристрої. Деякі типи проксі-серверів (наприклад, проксі-сервери HTTPS) можуть використовувати SSL / TLS для захисту з'єднань, але це не є обов'язковою функцією і зазвичай обмежується лише веб-трафіком.

Важливою відмінністю є механізми аутентифікації та управління

сеансами. У той час як проксі-сервери зазвичай використовують прості методи аутентифікації, такі як базова аутентифікація HTTP і фільтрація IP, VPN реалізують розширені протоколи шифрування, а Cryproxy не підтримує функції автоматичного відновлення з'єднання і відпрацювання відмови, які є стандартними для рішень VPN. Крім того, проксі-сервери часто мають обмежену пропускну здатність і відсутність оптимізації на рівні мережі, що може призвести до значних затримок в обробці запитів.

Крім того, проксі-сервер не надає важливих функцій безпеки, таких як захист від витоків DNS-запитів, блокування шкідливого трафіку на мережевому рівні та ізоляція мережевого стека користувача.

Але з іншої сторони, коли використовувати парсер, то потрібно для мови програмування і пз зазначати проксі сервер, про що детальніше в розділі 2.2.

2.2. Технології автоматизованого підключення до Tor

Щоб використовувати ПЗ для автоматичного збору даних з сайту onion, потрібно встановити з'єднання через проксі-сервер, що працює в мережі TOR. Це досягається за допомогою протоколу SOCKS5, з додатковим модулем для проксі-з'єднання. Алгоритм приведений на рисунку 2.5. Протокол SOCKS5 запускається, коли клієнт надсилає перший запит на підключення до сервера SOCKS5. Клієнт надсилає на сервер список усіх підтримуваних методів аутентифікації.

Після отримання цього списку сервер перевіряє рекомендований метод. Якщо сервер не приймає жодного із запропонованих методів, з'єднання негайно закривається. Якщо буде знайдено прийнятний метод, сервер вибере 1 з них для подальшої автентифікації. Потім сам процес аутентифікації виконується з використанням обраного методу. Якщо автентифікація не вдасться, з'єднання буде закрито. Після успішної аутентифікації декомунізується повне з'єднання між клієнтом і сервером. Згодом клієнт може надсилати запити через встановлене з’єднання SOCKS5.

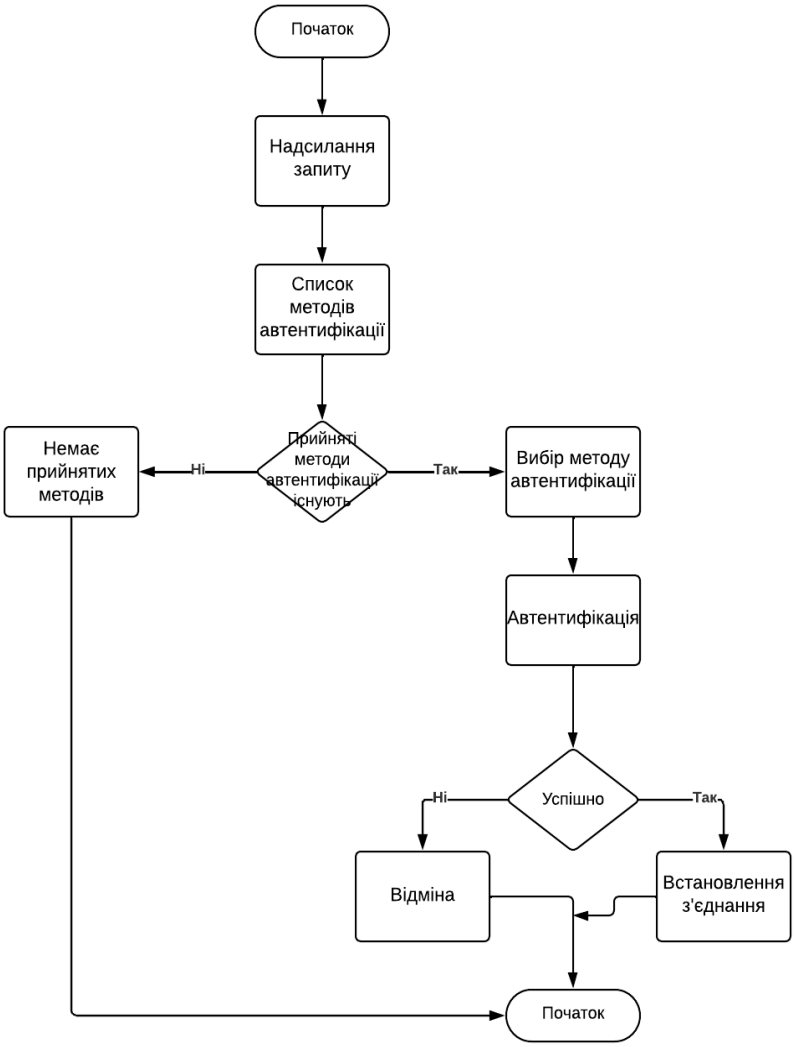


Рисунок 2.5 – Схема алгоритму роботи проксі – сервера SOCKS5

Для парсингу потрібно застосовувати порт 9050 - порт для роботи з Tor. Для більш ефективного збору даних на форумах Tor рекомендовано використовувати додаткові інструменти для управління підключеннями через Tor, такі як бібліотеки stem.

Відмінністю між звичайним підключенням до Tor людиною і автоматизованим є те, що людина виконує ряд умов, зазначених в розділі 2.1. для підключення до тор, коли в свою чергу для автоматизації потрібно виконати ряд інших умов.

Якщо говорити про автоматизоване ПЗ, то через API бібліотеки скрипт налаштовує параметри з'єднання - може вказати конкретні вузли для побудови каналу, задати країни, через які буде йти трафік, встановити таймери для зміни маршруту. Після запуску скрипт може автоматично створювати нові ланцюжки

з'єднань, міняти IP-адреси за розкладом, перевіряти статус з'єднання та перепідключатись при розірванні зв'язку. Все це відбувається без участі користувача - достатньо лише запустити скрипт, який розпочинається з TorController'a.

TorController програмно керує налаштуваннями та фоновими операціями TOR через STEM API. Основою маршрутизації в мережі Tor є те, що вузол входу діє як початкова точка входу в Мережу, а система вузлів, що шифрує початковий трафік, називається проміжний перехід, який маршрутизує зашифрований трафік між вузлами, забезпечуючи додатковий рівень шифрування та анонімізації. Вихідний вузол-це кінцева точка, в якій трафік залишає мережу Tor і надсилається цільовому джерелу.

Параметри якості обслуговування (QoS) дозволяють оптимізувати продуктивність з'єднання за рахунок визначення пріоритетів трафіку, управління пропускною здатністю і затримками.

Chemes - це криптографічно безпечні маршрути через вибрані мережеві вузли, які динамічно створюються та модифікуються для забезпечення анонімності. Система управління часом базується на конструкції, подібній до cron - можна виконувати обертання ланцюга, зміну IP-адреси та інші операції за певним графіком, гарантуючи, що маршрут регулярно оновлюється, а статус з'єднання оновлюється.

STEM постійно відстежує показники продуктивності, такі як затримка з'єднання, пропускна здатність каналу та стан вузла. При виявленні спотворення параметрів ланцюжок автоматично переналаштовується шляхом вибору найкращого вузла на основі останніх даних про стан мережі. Збір та аналіз телеметрії можуть оптимізувати маршрутизацію та забезпечити стабільність. Криптозахист в свою чергу забезпечується багаторівневим шифруванням трафіку на кожному вузлі з використанням алгоритмів асиметричного шифрування і протоколів узгодження ключів. Це забезпечує конфіденційність і цілісність даних, що передаються по мережі. В таблиці 2.1. буде представлено різницю між асинхронністю та синхронністю.

Таблиця 2.1. – Різниця між асинхронним та синхронним шифруванням.

Характеристика	Синхронне шифрування	Асинхронне шифрування
Ключі	Один спільний ключ для шифрування і дешифрування	Пара ключів: публічний для шифрування, приватний для дешифрування
Швидкість роботи	Висока швидкість обробки даних	Нижча швидкість через складніші обчислення
Безпека передачі ключів	Потребує захищеного каналу для обміну ключами	Не потребує захищеного каналу для публічного ключа
Використання в Tor	Для швидкої передачі даних після встановлення з'єднання	Для безпечного встановлення з'єднання та обміну сеансовими ключами
Масштабованість	Складно масштабується через необхідність обміну ключами	Легко масштабується, кожен вузол має власну пару ключів
Навантаження	Менше навантаження на процесор	Більше навантаження на процесор
Вразливість	Компрометація ключа впливає на все з'єднання	Компрометація одного вузла не впливає на інші
Довжина ключа	Коротші ключі (128-256 біт)	Довші ключі (2048-4096 біт)

При встановленні з'єднання кожен вузол мережі Tor використовує свою пару ключів - публічний та приватний. Публічний ключ відкрито поширюється в мережі, в той час як приватний надійно зберігається тільки на вузлі. Коли дані проходять через ланцюжок вузлів, кожен шар шифрування накладається з використанням публічного ключа відповідного вузла. Розшифрувати ці дані може тільки той вузол, який має відповідний приватний ключ. Така архітектура забезпечує унікальну властивість - навіть якщо зловмисник перехопить зашифрований трафік, він не зможе його розшифрувати без доступу до приватних ключів всіх вузлів у ланцюжку. Більше того, кожен вузол в ланцюжку

знає тільки адресу попереднього та наступного вузла, але не має інформації про повний маршрут.

При цьому асиметричне шифрування дозволяє безпечно узгоджувати сеансові ключі для швидкого шифрування без необхідності попереднього обміну секретними даними. Це особливо важливо в розподіленій мережі, де вузли постійно змінюються та не мають прямого захищеного каналу зв'язку між собою.

2.3. Системи скрапінгу даних з дарквебу

Після підключення до ресурсу Dark Web, варто розробити алгоритм збирання даних. В цьому допоможе скрапінг. Скрапінг (Web Scraping) - “це метод автоматизованого збору даних з веб-сайтів.

Програма, яка виконує скрапінг, аналізує HTML-код сторінок, знаходить потрібні елементи та витягує з них інформацію” [55].

Першим етапом веб-скрапінгу є завантаження HTML-коду веб-сторінки. Веб-скрапер надсилає HTTP-запит до цільового URL-адреси, аналогічно тому, як це робить веб-браузер.

Сервер повертає HTML-документ, який містить структуру та контент веб-сторінки. Веб-скрапер отримує цей HTML-код для подальшого аналізу. У деяких випадках для доступу до певних веб-сторінок може знадобитися аутентифікація або додаткові заголовки запиту.

Веб-скрапер повинен забезпечити належне опрацювання файлів cookie, заголовків та параметрів запиту, щоб успішно завантажити цільову сторінку. На рисунку 2.6 проілюстровано різницю між звичайним сайтом та сайтом з авторизацією, але на основі звичайних сайтів, для розуміння того, що саме може трапитись, коли ПЗ намагатиметься достукатись до сайту з авторизацією.

Звісно, це на прикладі гітхабу, але суть зрозуміла – при знаходженні сторінки без доступу, на прикладі з одного з репозиторіїв можна побачити, що неавторизований користувач не може переглядати закритий репозиторій, в той час як авторизований може.

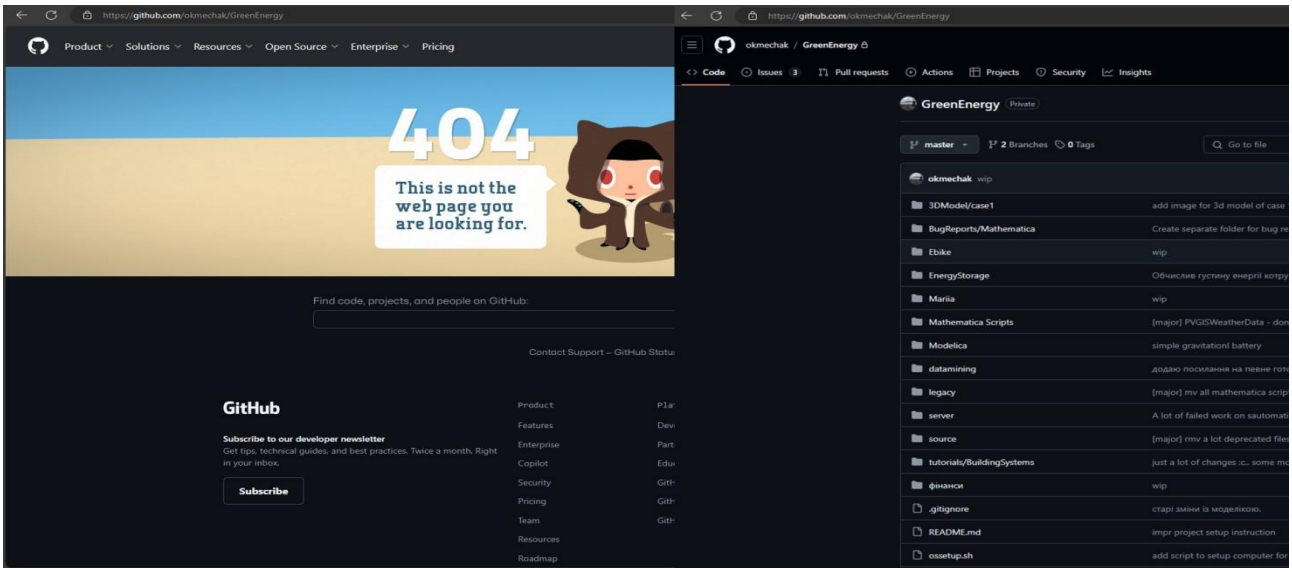


Рисунок 2.6 – Порівняння авторизованого і неавторизованого користувача на прикладі GitHub

Після успішного отримання HTML-коду сторінки, він передається парсеру для структурного аналізу та вилучення потрібних даних. HTML являє собою ієрархічний документ, де кожен елемент описується відповідним тегом. Парсер будує дерево об'єктної моделі документа (DOM), що дозволяє зручно обходити вузли і шукати необхідну інформацію. Для навігації по DOM використовуються селектори CSS або XPath-вирази. Вони дозволяють гнучко вказувати шляхи до елементів на основі їх тегів, класів, ідентифікаторів, атрибутів та взаємного розташування.

Наприклад, селектор CSS "div.product > h3.title" вибере всі заголовки h3 з класом "title", які знаходяться всередині блоків div з класом "product". XPath-вираз "//table[@id='prices']//tr[td[2]>10]" знайде всі рядки tr всередині таблиці з ідентифікатором "prices", в яких друга колонка td має значення більше 10. Окрім CSS та XPath, для пошуку та вилучення даних також можуть застосовуватись регулярні вирази. Вони особливо ефективні для пошуку фрагментів тексту, що відповідають певним шаблонам, наприклад, email-адресам, номерам телефонів, цінам тощо. На рисунку 2.7 зображено приклад того, що саме буде шукати запропоноване ПЗ на прикладі браузера Tor.

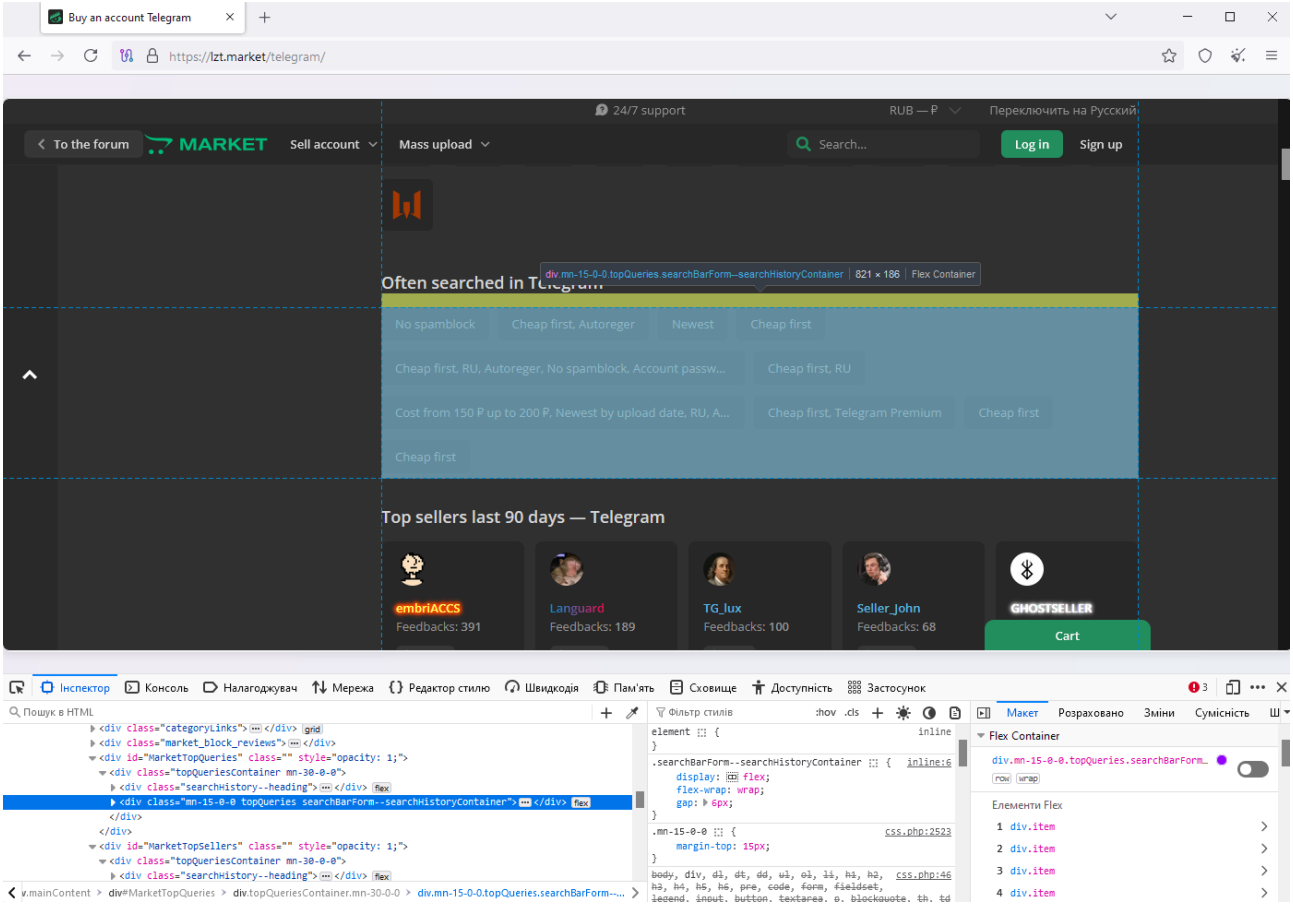


Рисунок 2.7 – Код елемента і основні вказівники для ПЗ

Веб-скрапінг ускладнюється, коли сайти використовують динамічне завантаження контенту за допомогою JavaScript. В таких випадках HTML-код, отриманий через звичайний HTTP-запит, може не містити повних даних, оскільки вони підвантажуються пізніше браузером. Для вирішення цієї проблеми застосовуються спеціальні інструменти, такі як Puppeteer або Selenium, які дозволяють імітувати повноцінну роботу браузера і виконувати JavaScript-код для отримання фінальної HTML-розмітки. Проте, це не є критичним пунктом для ПЗ, оскільки зазвичай TOR ресурси використовують примітивні візуальні елементи для простоти перенесення, швидкості переходу та створення лише потрібного функціоналу.

З іншої сторони, також відсутні і плюси складних систем, адже пагінація стає також неможливим елементом для таких сайтів, оскільки в дуже малій кількості для таких форумів пишуть бек-енд, ще й додають відкрите АПІ для можливості зручно користуватись сайтом.

На рисунку 2.8 відповідно зображено вихідний код одного з сайтів з

сегменту .onion.

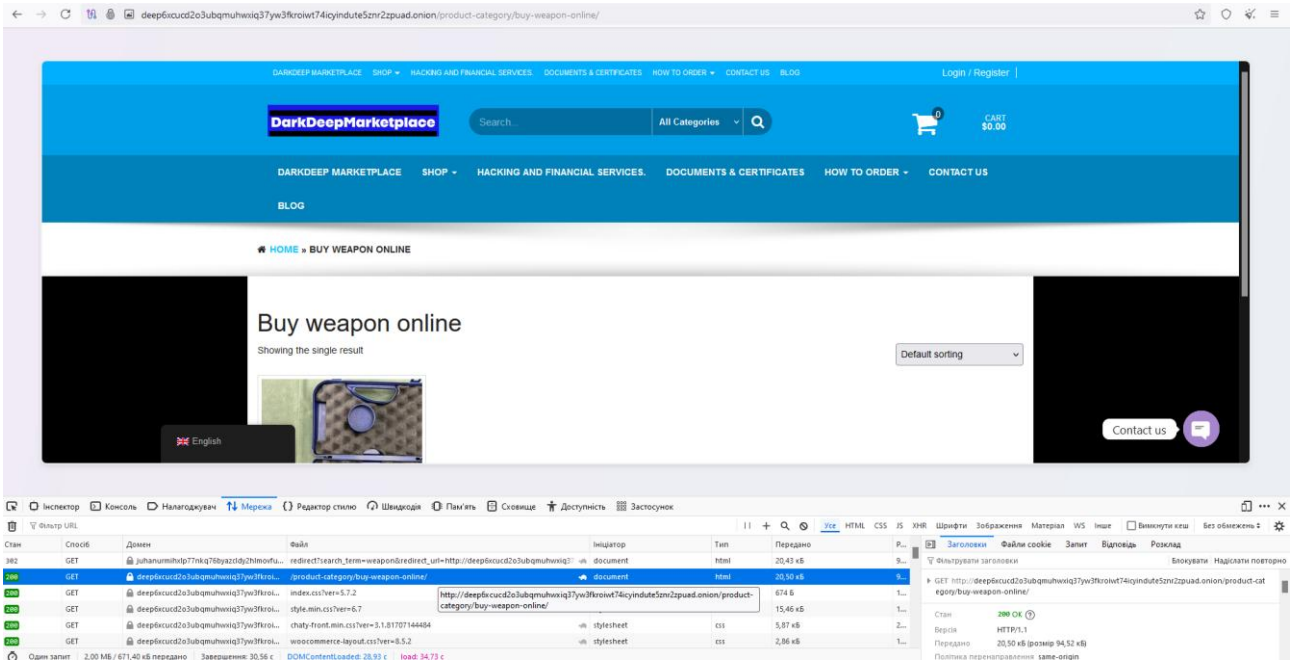


Рисунок 2.8 – Магазин продажу зброї (на основі .onion)

Як видно на рисунку, сама структура пошуковика не складається з якихось значних елементів, не підгружає скрипти, а сторінка зі зброєю в вкладці network відображається як окремий темплейт, який потрібно підгрузити в майбутньому. Тобто, бекенд частини на таких сайтах немає. На рисунку 2.9 зображено посилання записане через href, яке веде на інші елементи магазину.

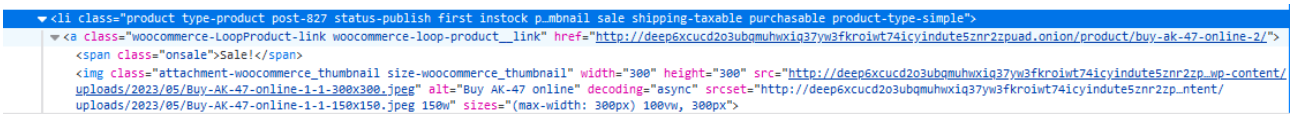


Рисунок 2.9 – Переведення на іншу сторінку магазину по продажу зброї

Для зручності роботи з отриманими даними, їх зазвичай зберігають у структурованих форматах, таких як CSV, JSON або XML. Це полегшує подальшу обробку, аналіз та інтеграцію з іншими системами. Також веб-скрапери часто взаємодіють з базами даних для довготривалого зберігання та ефективного доступу до зібраної інформації, проте для демонстрації можливості роботи достатньо буде CSV зі звичайним делімітером, наприклад, {;}. Оскільки в майбутньому можна буде під’єднати його до телеграм-боту або конвертувати csv

в файл xslx, що є більш розповсюдженим елементом перегляду інформації, ніж бази даних.

2.4. Алгоритм пошуку даних з джерел

“Формат значень, розділених комами (CSV), є одним з найпопулярніших форматів для зберігання та управління структурованими даними. Файл CSV-це текстовий файл, де дані організовані у вигляді таблиць, кожен рядок відповідає окремому запису, а кожне поле запису розділене роздільником. У більшості випадків в якості роздільника використовується кома (,) або крапка з комою (;), але в залежності від характеристик даних і вимог до обробки можуть використовуватися і інші символи” [56]. Структура файлу CSV передбачає, що кожен рядок містить однакову кількість полів, що відповідають певному атрибуту запису. Перший рядок файлу може містити заголовок, що описує назву поля або стовпця в таблиці. Наявність заголовка необов'язкова, але його використання значно полегшує розуміння та подальшу обробку структури даних. На рисунку 1.3. зображено приклад XSLX, який в відповідає майбутньому формату CSV.

При роботі з файлами CSV слід враховувати можливість пропуску або нульового значення в деяких полях. Пропуски можуть виникати, коли значення певного атрибута в записі відсутнє або невідоме. У цих випадках поле може залишатися порожнім або містити спеціальні символи (наприклад, "-" або "NaN"), що вказують на відсутність даних. Після заповнення csv даними, потрібно також їх діставати для зручного представлення. Перед тим як, розробити алгоритм, варто розглянути як сформована CSV на прикладі декількох значень для уявлення майбутнього парсингу даних.

На рисунку 2.10 зображено частину вмісту тестового CSV файлу, який в майбутньому буде підлягатись парсингу для збирання даних та відправку їх у програмне забезпечення для покращення розуміння та візуалізації даних, які були отримані в ході скрапінгу ресурсів Dark Web.

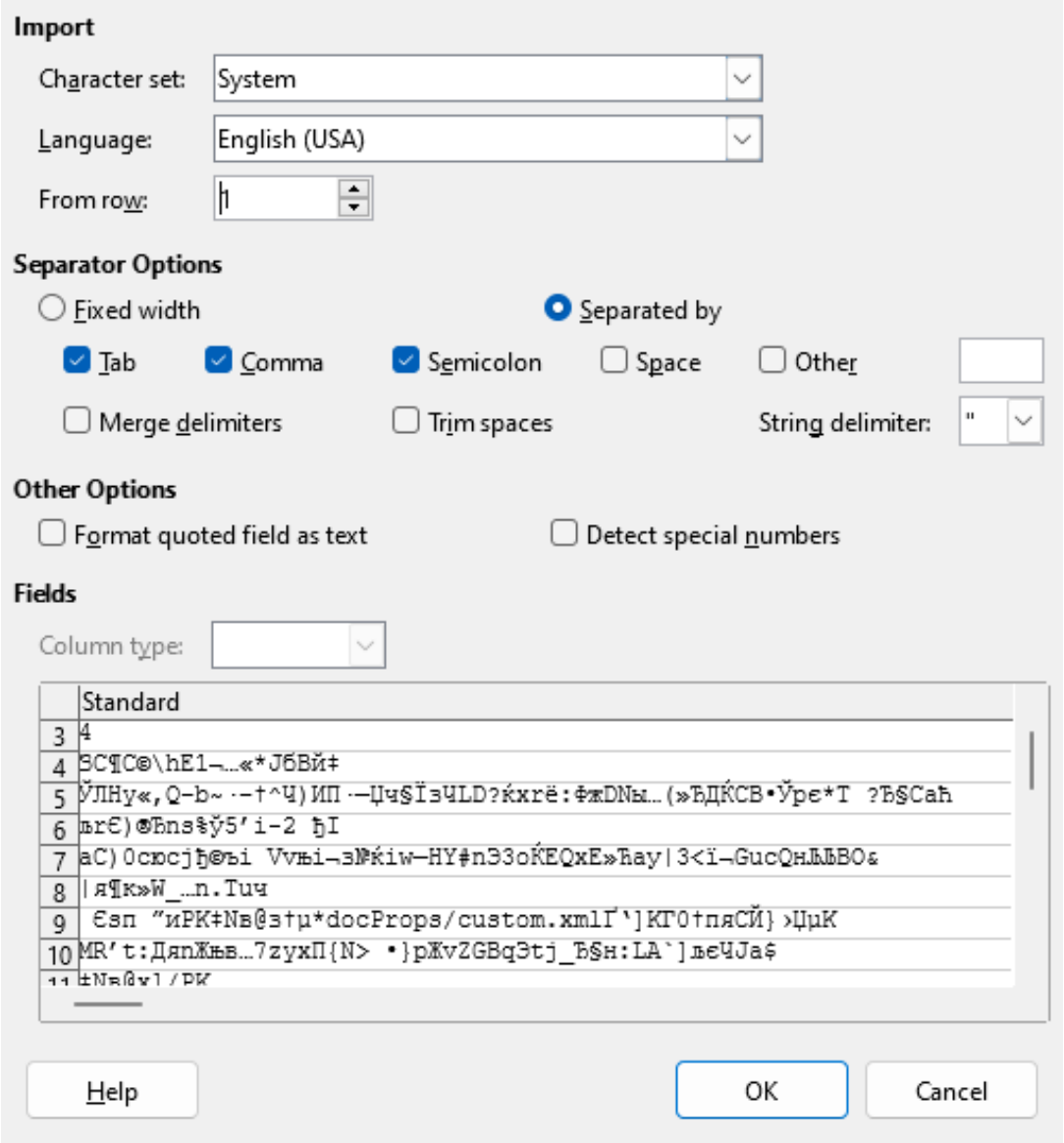


Рисунок 2.10 - Приклад вигляду тестового CSV

Відповідно, на рисунку 2.11 зображено алгоритм роботи парсера даних з CSV для представлення користувачеві даних, де на початку файл зчитується, делімітером розбивається на поля, які пізніше будуть парситись, опісля йде пошук по першому рядку, чи є таке поле, яке могло стягнути зі сайту ПЗ, якщо таке існує, відповідно по ньому буде відбуватись пошук у файлах CSV. Парсер аналізує кожен рядок файлу і розділяє його на окремі поля відповідно до заданого роздільника.

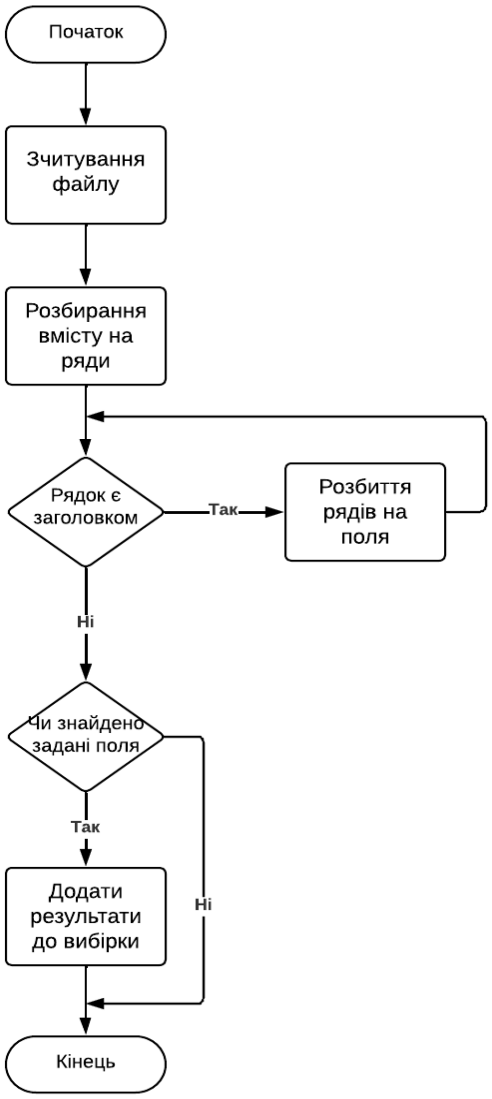


Рисунок 2.11 — Схема роботи алгоритму парсингу з CSV

Це дозволяє структурувати дані і забезпечити зручний доступ до кожного окремого значення. Після чого відбудеться формування звіту – збір всіх знайдених полів і відправка їх у UI частину ПЗ.

Висновки до розділу 2.

1. Проведено аналіз технологій анонімного доступу та безпеки, які широко використовуються в Dark Web для забезпечення конфіденційності та захисту користувачів. Зокрема, детально розглянуто протокол Tor (The Onion Router), який є основою анонімної мережі Dark Web. Досліджено принципи роботи Tor, включаючи багатошарове шифрування, цибулеву маршрутизацію та

використання вузлів-ретрансляторів для приховування реального місцезнаходження та ідентичності користувачів. Також проаналізовано використання віртуальних приватних мереж (VPN) та проксі-серверів як додаткових інструментів для забезпечення анонімності в Dark Web. Розглянуто криптографічні інструменти, такі як асиметричне шифрування, цифрові підписи та протоколи обміну ключами, які використовуються для захисту комунікацій та транзакцій в Dark Web.

2. Проаналізовано технології автоматизованого підключення до мережі Tor, які дозволяють спростити процес налаштування та управління з'єднаннями з Dark Web'ом. Розглянуто програмні засоби та бібліотеки, такі як Tor Controller Protocol (STEM) та Tor Browser Selenium, які надають можливість автоматизувати процес встановлення з'єднання з Tor, керувати налаштуваннями проксі, змінювати IP-адреси та виконувати інші операції. Досліджено методи інтеграції цих технологій у програмні додатки для забезпечення безпечного та анонімного доступу до ресурсів Dark Web.

3. Розглянуто системи скрапінгу даних з Dark Web'у, які дозволяють автоматизувати процес збору інформації з прихованих веб-ресурсів. Проаналізовано основні підходи та інструменти для скрапінгу даних, такі як використання парсерів HTML/XML, емуляція дій користувача за допомогою headless-браузерів (наприклад, Puppeteer або Selenium), а також обробка динамічного контенту та подолання антискрапінгових механізмів. Досліджено особливості застосування цих технологій в контексті Dark Web, де додатково необхідно враховувати анонімність, безпеку та обхід специфічних обмежень.

4. Описано алгоритм пошуку даних з джерел, зібраних в Dark Web, на основі формату CSV. Запропонований алгоритм дозволяє ефективно обробляти та аналізувати великі обсяги структурованих даних, отриманих в результаті скрапінгу дарквебу. Алгоритм включає етапи попередньої обробки даних, такі як очищення, нормалізація та видалення дублікатів, а також реалізує ефективні методи індексації та пошуку інформації на основі заданих критеріїв. Описано процес вилучення релевантних даних з CSV-файлів, їх структурування та збереження в зручному для подальшого аналізу форматі.

3 РОЗРОБКА СИСТЕМИ МОНІТОРИНГУ DARK WEB - РЕСУРСІВ

3.1. Реалізація UI для ПЗ

Для реалізації системи було взято за основу мову Python та бібліотеку PyQT5, яка дозволяє створювати інтерфейси для програмного забезпечення. PyQt5 - це набір Python-пов'язаних прив'язок бібліотеки Qt, однієї з найпопулярніших та найпотужніших крос-платформенних фреймворків для створення графічних інтерфейсів користувача (GUI). PyQt5 дозволяє розробникам використовувати всі можливості Qt у своїх Python-додатках. Ця бібліотека надає широкі можливості для створення повнофункціональних графічних програм з використанням різноманітних віджетів, таких як вікна, кнопки, меню, панелі інструментів, таблиці, графіки тощо. Розробники також можуть використовувати PyQt5 для створення додатків з багатопотоковістю та асинхронними операціями, а також кросплатформених програм, що працюють на Windows, macOS, Linux, Android та iOS.

Крім того, PyQt5 має високорівневе API, що забезпечує швидку розробку складних GUI-додатків, дозволяє створювати користувацькі віджети та елементи управління, а також надає підтримку анімацій, 2D/3D графіки, WebKit та інших сучасних можливостей. Встановити PyQt5 можна за допомогою pip:

```
pip install PyQt5
```

Після встановлення можна імпортувати PyQt5 у Python-коді та починати створювати вікна, додавати віджети, обробляти події та розробляти повноцінні GUI-додатки.

Також хорошим плюсом PyQT5 є те, що можна застосовувати елементи з таблиці каскадів CSS, і ще вона є ідеальним вибором, оскільки проста в освоєні і застосуванні, де все прописується блочною системою і стандартні функції пайтона можна підв’язати до кнопки в один аргумент кнопки чи лейбла, тобто ніякі додаткові зв’язні методи між UI та кодом не знадобляться.

Стосовно візуалу, перед тим в figma було створено приблизний візуал для майбутнього його відтворення через PyQt5. На рисунку 3.1. зображено приблизний дизайн веб-скрапера для ресурсів дакнету

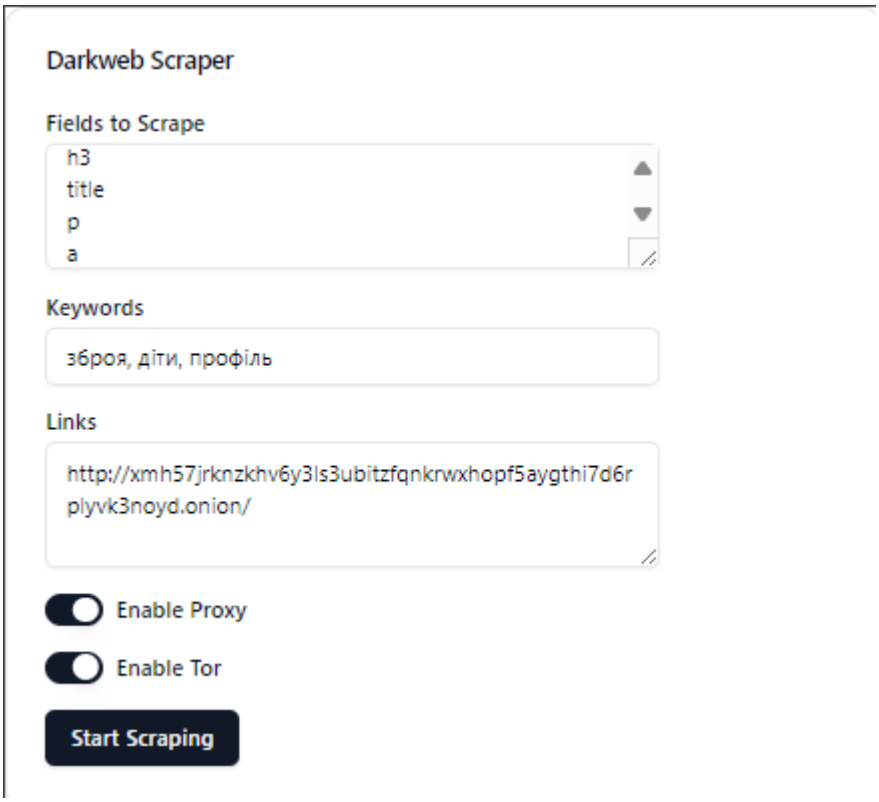


Рисунок 3.1 – Веб Скрапер на етапі проектування

На рисунку зображено декілька полів, де перший Fields to Scrape відповідає за те, які теги на сайтах будуть застосовуватись для пошуку. Має бути функція, яка дозволяє шукати по всім тегам - __all__. Другим має бути поле Keywords, власне, по якому і буде відбуватись пошук. Це ключові слова, які мають зустрітись в тегах, щоб в подальшому з батьківського елементу можна було витягнути посилання, елементи і сформувати CSV.

Окремими полями було добавлено Proxy та CSV. Для проксі після перегортання світча було добавлено новий лейбл, який з’являється після перемикання його у статус True. Це дозволить користувачеві вводити власні проксі-сервера, якщо старий, наприклад, перестав працювати. На рисунку 3.2. зображено нове поле.

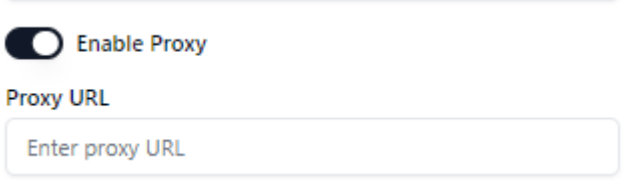


Рисунок 3.2 – Поле для внесення адреси проксі

Для реалізації було встановлено бібліотеку і виготовлено приблизний шаблон майбутнього UI. На рисунку 3.3 представлено як виглядає візуал без застосування CSS, але перенесений вже в код.

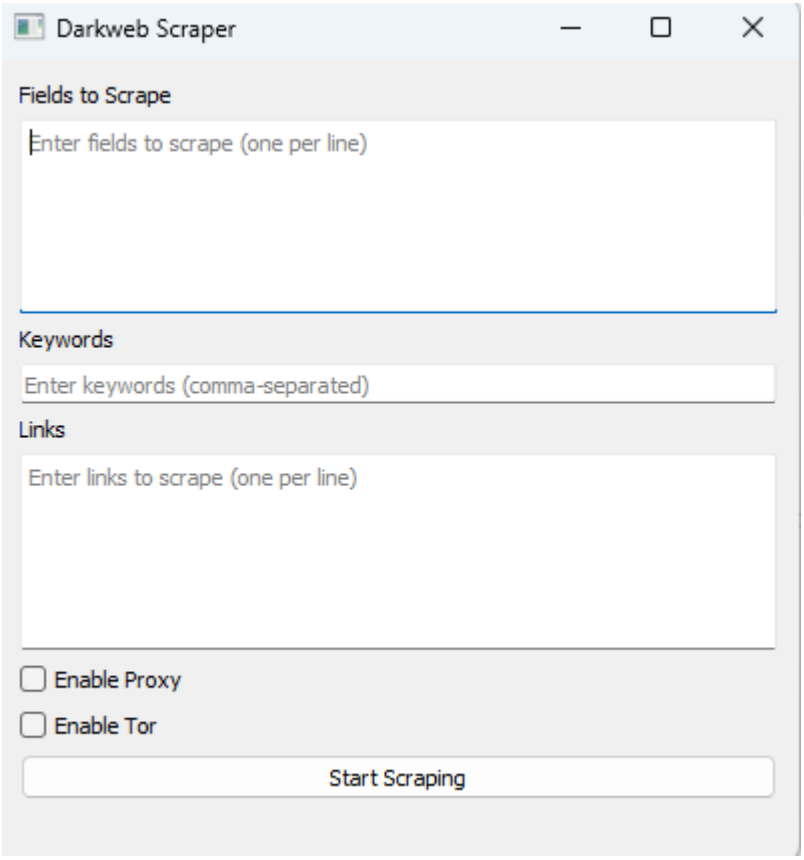


Рисунок 3.3 – Проектування головного вікна програми

Також було прописано логіку, яка дозволяє увімкнути поле вводу проксі, якщо чекбокс Proxy має галочку:

```
def toggle_proxy(self, state):
    if state == 2:
        self.proxy_input.setVisible(True)
    else:
        self.proxy_input.setVisible(False)
```

Результат вибору опції «Enable Proxy», зображено на рисунку 3.4.

Рисунок 3.4 – Проектування полів введення

Накладання стилів відбувається шляхом приписування схожих до CSS команд для елементів. Прикладом може послужити приклад коду, зазначений нижче, що буде відображено також на рисунку 3.5:

```
self.setStyleSheet("""
        QWidget {
            background-color: #FFFFFF;
            color: #1E1E1E;
            font-family: Arial, sans-serif;          }
        QLabel {
            font-size: 14px;
            margin-bottom: 5px;          }
```

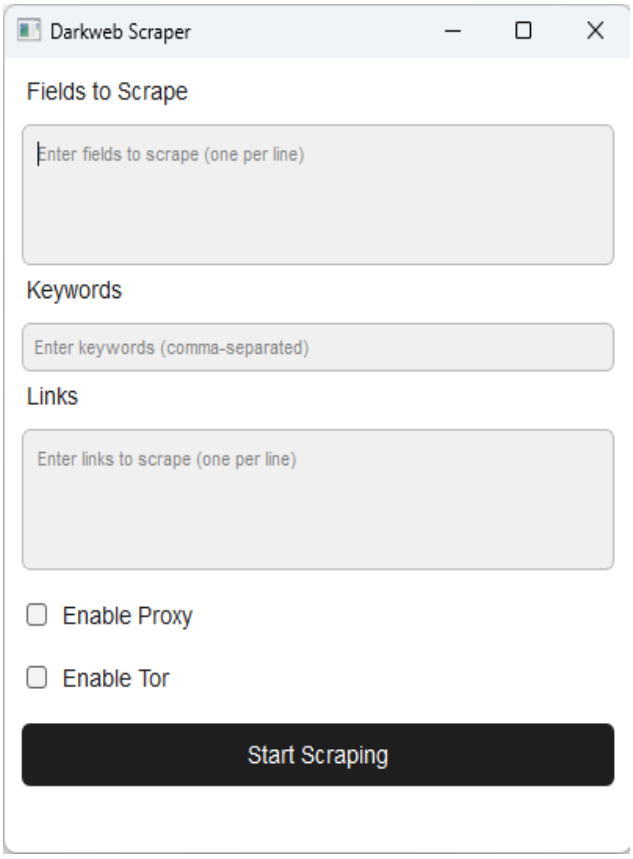


Рисунок 3.5 – Написання CSS для UI

3.2. Реалізація системи моніторингу

Перш за все, потрібно визначити структуру проекту, окрім того файлу, який містить назву UI.py. Майбутня структура проекту має виглядати наступним чином:

```
darkweb_monitor/
├── monitor.py
├── UI.py
├── scraper.py
├── notifier.py
├── config.py
└── requirements.txt
```

Якщо розповідати про кожен файл, то перший в свою чергу про config.py, де буде записуватись основний конфіг, який підгружатиметься, якщо людина не вводить ніякі дані для перевірки. Тобто, config.py містить загальну інформацію про сайти, які потрібно моніторити, стандартний проксі і інтервал, з яким програма працюватиме.

```
MONITORING_SITES = [
    "http://darkwebsite1.onion",
    "http://darkwebsite2.onion"
]
```

Великими буквами виділено MONITORING_SITES, оскільки це змінні, які є константами та не будуть змінюватись ніколи. Також варто зразу встановити дефолтний кофніг, який буде застосовуватись:

```
TOR_PROXY = "socks5h://127.0.0.1:9050"
```

І відповідно, CHECK_INTERVAL відповідає за час спрацювання коду:

CHECK_INTERVAL = 3600 – виставлений на одну годину.

Таким чином, оскільки з’явились два файли, які в майбутньому потрібно буде об'єднати, потрібно також прописати файл __init__.py, який вкаже, що всі файли з поточної теки - модульні файли, які в майбутньому можна буде

імпортувати в основний документ через:

from config import TOR_PROXY

Основні компоненти системи включають парсинг сайтів за допомогою бібліотек для веб-скрапінгу, пошук нової або зміненої інформації, сповіщення про зміни через електронну пошту або інші методи та інтерфейс для налаштувань і моніторингу.

Для реалізації системи потрібно використовувати бібліотеки, такі як requests для HTTP-запитів, BeautifulSoup з бібліотеки bs4 для парсингу HTML, schedule для налаштування періодичних перевірок та stem для роботи з Tor (якщо моніторинг здійснюється через Tor).

Структура проекту включатиме декілька основних файлів. config.py містить конфігурацію для налаштувань, таких як сайти для моніторингу, налаштування Tor і пошти. scraper.py буде відповідати за отримання та парсинг контенту сторінок DarkWeb. У файлі notifier.py буде реалізовано надсилання сповіщень, наприклад, через електронну пошту. Основна логіка моніторингу буде зібрана у файлі monitor.py.

В конфігураційному файлі config.py можна зазначити список сайтів для моніторингу, налаштування Tor (якщо використовується), налаштування для сповіщень, а також інтервал часу для перевірки сайтів. У файлі scraper.py реалізовано функцію отримання контенту сторінки через Tor та парсинг HTML для збору необхідної інформації. Функція get_page_content отримує сторінку через Tor, використовуючи налаштування проксі, а функція parse_page парсить контент сторінки, наприклад, для пошуку заголовків новин або іншої важливої інформації.

У файлі notifier.py реалізована функція для надсилання електронної пошти, яка використовує бібліотеку smtplib для підключення до SMTP-сервера та надсилання повідомлень. Основна програма для моніторингу працює у файлі monitor.py. Вона періодично перевіряє сайти зі списку, отримує їх контент за допомогою функції get_page_content і парсить його за допомогою parse_page. Якщо вміст сайту змінюється, програма надсилає сповіщення на електронну пошту. Всі перевірки виконуються з певним інтервалом часу, який можна

налаштувати. Також буде підключенно і UI.py, який відповідно підгружатиме основний код і об’єднуватиме методи з кнопками та лейблами.

В подальшому, щоб не було помилок, які були викликані на рисунку 3.6. також сформовано requirements.txt, куди входитимуть всі майбутні модулі, пов’язані з проектом.

```
    import schedule
ModuleNotFoundError: No module named 'schedule'
```

Рисунок 3.6. Відсутній модульний пакет.

Requests - бібліотека для виконання HTTP-запитів. Вона дозволяє вам легко взаємодіяти з веб-сайтами та API, отримуючи HTML-контент, файли або дані у форматі JSON, а також відправляючи дані на сервер.

beautifulsoup4 - бібліотека для парсингу HTML і XML документів. Вона дозволяє витягувати та обробляти дані з веб-сторінок, наприклад, знаходити елементи за тегами, класами, ідентифікаторами та іншими атрибутами.

stem - бібліотека для взаємодії з Tor (мережею анонімності). Вона дозволяє підключати ваші запити до Tor, щоб маскувати вашу реальну IP-адресу, а також дає можливість змінювати ідентифікатори мережі для досягнення анонімності.

Після прописування команди pip install -r requirements.txt вікно, яке вказує, які модулі були підвантажені, як показано на рисунку 3.7.

```
Collecting stem
  Downloading stem-1.8.2.tar.gz (2.9 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.9/2.9 MB 8.3 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting schedule
```

Рисунок 3.7 – Встановлення необхідних модулів

Також було виявлено помилку, яка вказує на неможливість з’єднатись з проксі-сервером. Помилка stem.SocketError: [WinError 10061] No connection could be made because the target machine actively refused it вказує на те, що ваш запит до Tor-сервера не вдалося здійснити, оскільки підключення було відхилено.

Тому, було добавлено також код, який використовує порт 9051, це порт контролю, після чого, з 2 разу програма виконалась, що зображено на рисунку 3.8.

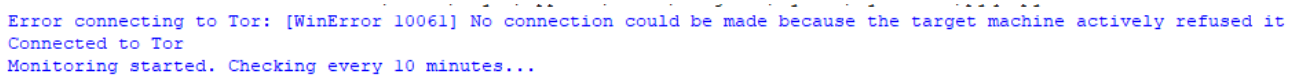


Рисунок 3.8 – Підключення до Tor через Python

Проте з’являється 2 помилка, яка зображена на рисунку 3.9.

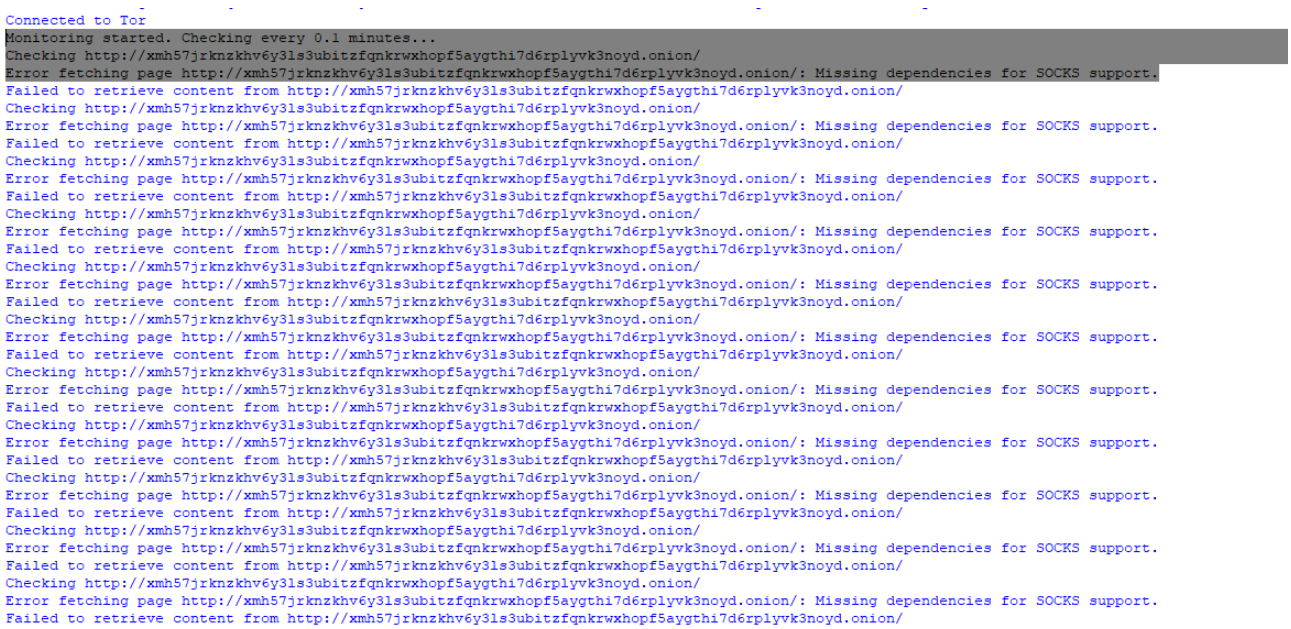


Рисунок 3.9 – Відсутність SOCKS модуля для сайту .onion

Це виправляється переналаштуванням проксі-серверів і обов’язковим ввімкненням Тору. Після того без headless режиму вмикається посилання, яке вказане в коді, а власне: http://xmh57jrknzkhv6y3ls3ubitzfqnkrwxhopf5aygthi7d6rplyvk3noyd.onion/cgi-bin/omega/omega” - посилання на torch. Популярний пошуковик з рекламою в сегменті .onion. Відповідно, результат хорошої роботи зображений на рисунку 3.10.

Рисунок 3.10 – Робота python без headless режиму

Headless режим дозволяє вмикати tor без його графічної оболонки, що зменшує навантаження системи.

Після правильного налаштування компонентів моніторингу, потрібно, також зробити сповіщення, якщо система знайшла щось підозріле. Найпростішим методом буде відправляти сповіщення на пошту, оскільки python дозволяє по SMTP відправляти дані з пошти на власну ж пошту. Для цього, був встановлений модуль, що продемонстрований на рисунку 3.6., а також написаний скрипт зі змінними, де зберігається інформація про протокол передачі, порт передачі і відповідно відправника, отримувача і пароль:

EMAIL_SENDER = myemail@gmail.com'

EMAIL_PASSWORD = 12345678

EMAIL_RECEIVER = 'myemail_2@gmail.com'

SMTP_SERVER = 'smtp.gmail.com'

SMTP_PORT = 587

А також реалізовано простий метод, який дозволяє відправляти збережені логи через захищене з’єднання:

```
def send_email(subject, body):
    msg = MIMEMultipart()
    msg['From'] = EMAIL_SENDER
    msg['To'] = EMAIL_RECEIVER
    msg['Subject'] = subject
    msg.attach(MIMEText(body, 'plain'))
    try:
        with smtplib.SMTP(SMTP_SERVER, SMTP_PORT) as server:
            server.starttls()  # Захищене з'єднання
```

```
            server.login(EMAIL_SENDER, EMAIL_PASSWORD)
            server.sendmail(EMAIL_SENDER, EMAIL_RECEIVER,
msg.as_string())
            print(f"Email sent to {EMAIL_RECEIVER}")
        except Exception as e:
            print(f"Error sending email: {e}")
```

Функція send_email відповідає за надсилання електронних листів через SMTP. Спочатку створюється об'єкт листа за допомогою MIMEMultipart, що дозволяє додавати кілька частин до листа, таких як текст та вкладення. Встановлюються параметри листа, зокрема адреси відправника та отримувача, а також тема листа. Текстовий вміст листа додається за допомогою MIMEText у форматі простого тексту. Далі функція підключається до SMTP-сервера через smtplib.SMTP, використовуючи зазначені сервер і порт. З'єднання встановлюється через захищене з'єднання TLS за допомогою методу starttls(). Потім виконується авторизація на сервері за допомогою методу login, де вказуються облікові дані відправника. Після цього за допомогою методу sendmail відправляється лист. Якщо виникає помилка при підключенні або відправці, вона обробляється та виводиться відповідне повідомлення про помилку.

Результат дії можна побачити на рисунку 3.11.

Рисунок 3.11 – Сповіщення моніторингу на gmail

Після чого, до надсилання було написано метод, який хапає csv з папки з датою, про що детальніше в пункті 3.3 і надсилає його як вкладений файл через функцію:

```
if attachment_path and os.path.isfile(attachment_path):
    filename = os.path.basename(attachment_path)
    with open(attachment_path, 'rb') as file:
      file_data = file.read()
      file_type = attachment_path.split('.')[-1]
```

file_type визначається окремо, оскільки сам файл передається закодованим в base64, відповідно тип файлу кодувати не треба, того його виділяють окремо. Проблем з додатковими крапками в назві файлу немає, тому що відбувається пошук по крапці з останнього символу, а не першого, а це завжди буде крапка, яка відділяє файл.

3.3. Реалізація компонентів збору та аналізу даних

В рядку from bs4 import BeautifulSoup імпортується клас BeautifulSoup з бібліотеки bs4, який дозволяє ефективно аналізувати та парсити HTML і XML документи. Це необхідно, оскільки без цього імпорту не можна буде створити об'єкти BeautifulSoup і використовувати їх для навігації по HTML-коду.

```
html_content = "<html><head><title>Test Page</title></head><body><h1>Welcome to the website</h1></body></html>"
```

Задається змінна html_content, що містить рядок з HTML-кодом. Це є вихідним документом, який потрібно обробити. В цьому випадку, цей HTML складається з базових тегів: <html>, <head>, <title> та <body>. Зазвичай, HTML контент отримується з реальної веб-сторінки за допомогою запиту через бібліотеку типу requests, але для демонстрації використовуємо статичний рядок.

Після цього створюється об'єкт BeautifulSoup за допомогою наступного рядка:

```
soup = BeautifulSoup(html_content, 'html.parser')
```

Цей рядок аналізує HTML-код, переданий у html_content, і перетворює його у зручну для роботи структуру — дерево об'єктів. Це дерево дозволяє зручно звертатися до окремих елементів документа і витягувати необхідну інформацію, наприклад, текст із тегів або атрибути.

Рядок soup = BeautifulSoup(xml_content, 'xml') використовується для створення об'єкта BeautifulSoup, який перетворює XML-документ у деревоподібну структуру даних. Це дозволяє легко переміщатися по елементах XML і знаходити потрібну інформацію, використовуючи інтуїтивно зрозумілі методи та функції. На відміну від HTML, XML є більш структурованим і суворим у форматуванні. Це означає, що кожен відкритий тег повинен мати відповідний закритий тег, а всі елементи повинні бути вкладені правильно. Використання BeautifulSoup для обробки XML допомагає швидко витягувати дані, що містяться у вкладених тегах або атрибутах. В коді воно виглядає наступним чином:

```
from bs4 import BeautifulSoup
# Імпортуємо клас BeautifulSoup із бібліотеки bs4 для аналізу HTML/XML документів
html_content = "<html><head><title>Test Page</title></head><body><h1>Welcome to the website</h1><p class='info'>This is a sample paragraph.</p></body></html>"
# Створюємо об'єкт BeautifulSoup для аналізу HTML
soup_html = BeautifulSoup(html_content, 'html.parser')
# Знаходимо елемент <h1> і отримуємо його текст
title = soup_html.find('h1').text
print("HTML Заголовок:", title)
# Знаходимо елемент <p> з класом 'info' і отримуємо його текст
paragraph = soup_html.find('p', class_='info').text
print("HTML Параграф:", paragraph)
```

Рядок 'title = soup.find('h1', class_='article-title')' використовує метод find() для пошуку першого елемента на сторінці, який відповідає заданим критеріям - тег h1 з класом 'article-title'. Даний метод дуже гнучкий і дозволяє шукати елементи за різними атрибутами, включаючи теги, класи, ідентифікатори та інші HTML-атрибути. Після знаходження потрібного елемента з нього можна отримати текст, атрибути або вкладені елементи. Рядок title = soup.find('h1', class_='article-title') використовує метод find() об'єкта BeautifulSoup для пошуку першого елемента на сторінці, який відповідає заданим критеріям. В даному

випадку шукається перший тег <h1>, у якому є клас 'article-title'. Метод find() повертає перший знайдений елемент, який відповідає умовам пошуку. Якщо такий елемент не знайдений, метод поверне значення None.

Метод find() є дуже гнучким, оскільки дозволяє шукати елементи за різними атрибутами, такими як тег (наприклад, 'h1'), клас (через параметр class_), ідентифікатор (через параметр id), атрибути або навіть текстові значення. Це дає можливість виконувати точний пошук та знаходити необхідні елементи на сторінці. Після того, як елемент знайдений, з нього можна отримати текст або інші атрибути. Наприклад, можна викликати title.text, щоб отримати текст, який міститься в цьому елементі, або title['href'], щоб отримати значення атрибута href, якщо елемент є посиланням.

Рядок 'content_div = soup.find('div', class_='article-content')' шукає основний контейнер статті на сторінці. Знайдений div елемент містить весь текстовий контент статті. З цього контейнера можна далі витягти всі параграфи тексту, використовуючи метод find_all(). Такий підхід дозволяє отримати структурований текст статті, розбитий на параграфи.

В таблиці 3.1. можна побачити порівняння методів і чому було використано саме метод find_all()

Таблиця 3.1 – Метод find - порівняння

Метод	Опис	Приклад	Повертає
find()	Шукає перший елемент, що відповідає заданим критеріям.	soup.find('a', class_='link')	Повертає перший знайдений елемент або None
find_all()	Шукає всі елементи, що відповідають критеріям.	soup.find_all('p', class_='text')	Повертає список усіх знайдених елементів
find_all() (без класу)	Шукає всі елементи за тегом без додаткових критеріїв.	soup.find_all('div')	Повертає список всіх на сторінці

Продовження таблиці 3.1

find() з атрибутами	Шукає перший елемент, що відповідає тегу та атрибуту.	soup.find('img', src='image.jpg')	Повертає перший знайдений елемент <img> з атрибутом src

Важливою частиною парсингу є обробка запитів до веб-сервера через рядок 'response = self.session.get(url, headers=headers, timeout=10)'. Цей код виконує HTTP GET запит до вказаного URL з користувацькими заголовками та обмеженням часу очікування. Використання користувацьких заголовків допомагає імітувати звичайний веб-браузер та уникнути блокування з боку сервера. Встановлення таймауту запобігає зависанню програми при повільних відповідях сервера. При роботі з списками новин використовується код 'news_blocks = soup.find_all('article', class_='news-item')' для знаходження всіх блоків новин на сторінці. Метод find_all() повертає список всіх елементів, що відповідають заданим критеріям пошуку. Кожен знайдений блок новини можна далі обробити для отримання заголовка, опису, посилання та інших деталей. Такий підхід дозволяє ефективно обробляти сторінки зі списками новин або інших однотипних елементів.

Для роботи з посиланнями використовується конструкція 'title_link = title_elem.find('a')' з подальшим отриманням URL через 'news_data['url'] = title_link.get('href')'. Пошук тега посилання всередині заголовка дозволяє отримати як текст заголовка, так і URL статті. Метод get() використовується для отримання значення конкретного атрибута HTML-елемента. Це корисно для збору URL-адрес, які потім можна використовувати для переходу на сторінки окремих статей.

Обробка помилок реалізується через конструкцію try-except: 'try: response = self.session.get(url) except requests.exceptions.RequestException as e:'. Такий підхід дозволяє коректно обробляти різні мережеві помилки, які можуть виникнути під час завантаження сторінок. При виникненні помилки код може спробувати повторити запит або логувати помилку для подальшого аналізу.

Також є функція, яка обробляє інші помилки:

```
def fetch_data(url):
    try:
        response = requests.get(url, timeout=5)
        response.raise_for_status()  # Перевірка на HTTP-помилки
        return response.text
    except requests.exceptions.ConnectionError:
        print("Помилка підключення. Сервер недоступний.")
    except requests.exceptions.Timeout:
        print("Час очікування вичерпано.")
    except requests.exceptions.HTTPError as http_err:
        print(f"HTTP помилка: {http_err}")
    except requests.exceptions.TooManyRedirects:
        print("Забагато перенаправлень. Спробуйте інший URL.")
    except requests.exceptions.SSLError:
        print("Помилка SSL-сертифіката.")
    except requests.exceptions.RequestException as e:
        print(f"Сталася помилка: {e}")
```

В таблиці 3.2. представлено список основних помилок, які необхідно було ловити.

Таблиця 3.2 – Види помилок

Тип винятку	Опис	Приклад використання	Дії при обробці
requests.exceptions.RequestException	Загальний базовий клас для всіх помилок бібліотеки requests.	except requests.exceptions. RequestException as e:	Логування помилки, повторення запиту або повідомлення користувачу

Продовження таблиці 3.2

requests.exceptions.Timeout	Виникає, якщо запит перевищує час очікування на відповідь від сервера.	except requests.exceptions.Timeout as e:

У таблиці 3.3 представлено менш значущі помилки, які можуть виникати рідше або за специфічних умов, проте їх теж варто враховувати під час розробки надійного парсера.

Таблиця 3.3 — Менш значущі помилки в requests

Тип виключення	Опис помилки	Можливі причини
requests.exceptions.URLRequired	Вказана некоректна або відсутня URL-адреса.	Відсутність протоколу (http/https) або використання некоректних символів в URL.
requests.exceptions.FileModeWarning	Попередження при використанні неправильного режиму файлу для запису/читання.	Використання режиму файлу, несумісного з операцією (наприклад, запис у файл, відкритий для читання).
requests.exceptions.InvalidSchema	Вказано невідомий або некоректний протокол у URL.	Використання невідомих схем, таких як ftp://, sftp://, що не підтримуються бібліотекою requests.

3.4. Формування файлів та їх парсинг

Після отримання даних сформовано CSV файл, який має назву today_date.csv. Цей файл містить мінімальну-необхідну інформацію про продаж зброї на сайті, який був перебраний парсером. Структура файлу CSV була розроблена з урахуванням необхідності зберігання мінімального, але адекватного набору даних для кожної зброї. При використанні формату UTF-8 спеціальні символи і кириличні символи відображаються правильно в різних операційних системах і пристроях. Це стає особливо важливим при роботі з описами продукції і технічними характеристиками зброї, включаючи певні символи і позначення.

Кожен рядок у файлі CSV представляє розташування окремих знарядь і містить ключові поля, необхідні для визначення та аналізу пропозиції. Ці поля включають унікальний ідентифікатор продукту, назву моделі, тип зброї, калібр, ціну, доступність, дату додавання оголошення та інші важливі функції. Якщо відкрити файл в блокноті з кодуванням UTF-8, можна візуально перевірити правильність збережених даних і наявність всіх необхідних полів.

При створенні файлу CSV особлива увага приділяється очищенню та нормалізації даних. Усі текстові поля виконують кроки для видалення непотрібних пробілів, спеціальних символів та тегів HTML, які можуть залишитися після аналізу. Цифри, такі як ціни та специфікації, перетворюються на стандартні формати, щоб забезпечити правильне сортування та фільтрацію даних.

Успішне виконання скрипта підтверджується наявністю у файлі повного набору записів з правильно введеними полями. Кожен запис містить унікальний ідентифікатор, який дозволяє відстежувати зміни в певному місці під час наступних оновлень даних. Це буде особливо важливо для відстеження змін цін і доступності товарів на ринку. Формат CSV був обраний через його універсальність та простоту використання. Такі файли відкриваються в різних програмах для роботи з електронними таблицями, таких як Microsoft Excel, Google Таблиці або спеціалізовані інструменти аналізу. Це забезпечує більшу

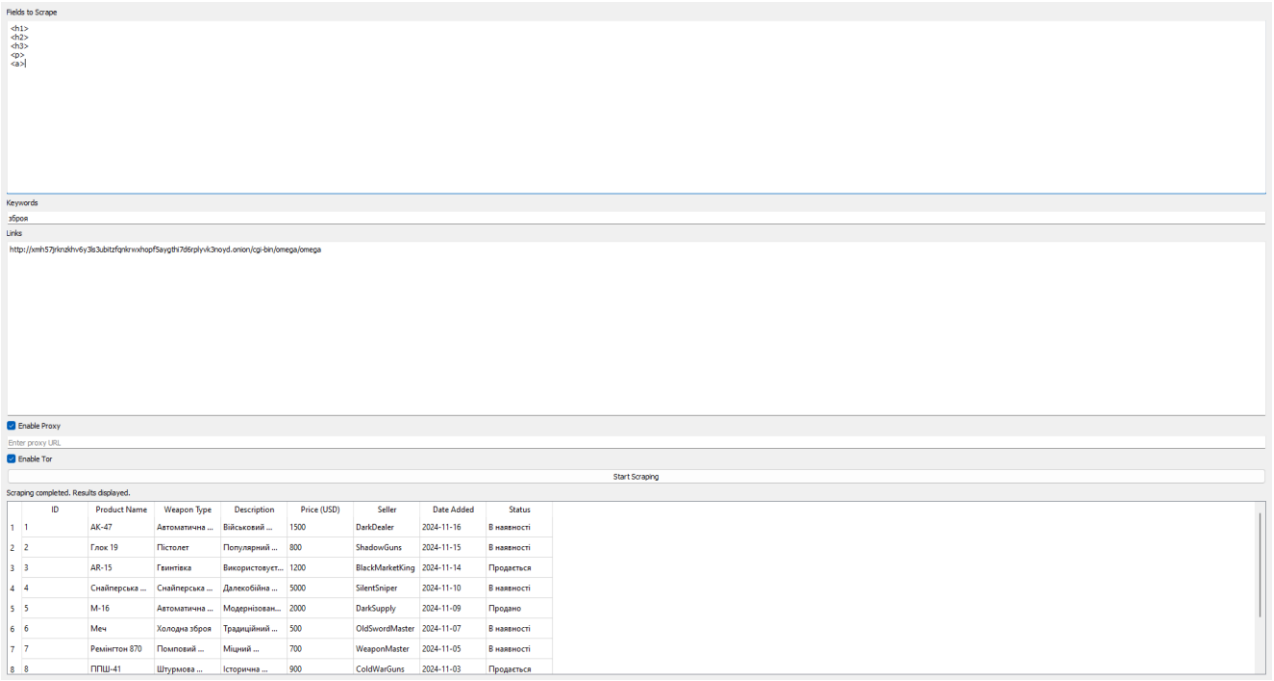
гнучкість в роботі.

На рисунку 3.12 зображено приклад відкритого файлу, спарешного з ресурсу дарквебу з правильними даними стосовно продажу зброї.

```
ID,Назва товару,Тип зброї,Опис,Ціна (USD),Продавець,Дата додавання,Статус
1,АК-47,Автоматична зброя,"Військовий автомат, бойова зброя, має вбудований магазин на 30 патронів.",1500,DarkDealer,2024-11-16,В наявності
2,Глок 19,Пістолет,"Популярний пістолет для самооборони. Висока точність, простота в експлуатації.",800,ShadowGuns,2024-11-15,В наявності
3,AR-15,Гвинтівка,"Використовується для полювання та самооборони. Можна налаштувати під різні патрони.",1200,BlackMarketKing,2024-11-14,Продається
4,Снайперська гвинтівка Barrett M82,Снайперська зброя,"Далекобійна гвинтівка калібру .50, для професіоналів.",5000,SilentSniper,2024-11-10,В наявності
5,М-16,Автоматична зброя,"Модернізована версія класичного AR-15 для військових.",2000,DarkSupply,2024-11-09,Продано
6,Меч,Холодна зброя,"Традиційний японський меч. Виготовлений вручну, висока якість сталі.",500,OldSwordMaster,2024-11-07,В наявності
7,Ремінгтон 870,Помповий дробовик,"Міцний помповий дробовик для самооборони та полювання.",700,WeaponMaster,2024-11-05,В наявності
8,ППШ-41,Штурмова гвинтівка,"Історична радянська штурмова зброя з патронами 7.62 мм. Оригінальна модель.",900,ColdWarGuns,2024-11-03,Продається
9,Uzi,Пістолет-кулемет,"Маленький і компактний пістолет-кулемет для близької дистанції.",1200,QuickGunDealer,2024-11-01,В наявності
10,Міні-міт,Міномет,"Міномет малого калібру для тактичних дій, з комплектом мін.",4000,FirePower,2024-10-30,Продано
```

Рисунок 3.12 – Отримані дані

На момент реалізації ще не було прописано стилі для програми, оскільки планувались її майбутні покращення. Проте, саме такий промпт, який зображений на рисунку 3.13. був заданий парсеру і отримана відповідь протягом хвилини.



	ID	Product Name	Weapon Type	Description	Price (USD)	Seller	Date Added	Status
1	1	AK-47	Автоматична ...	Військовий ...	1500	DarkDealer	2024-11-16	В наявності
2	2	Глок 19	Пістолет	Популярний ...	800	ShadowGuns	2024-11-15	В наявності
3	3	AR-15	Гвинтівка	Використовуєт...	1200	BlackMarketKing	2024-11-14	Продається
4	4	Снайперська ...	Снайперська ...	Далекобійна ...	5000	SilentSniper	2024-11-10	В наявності
5	5	М-16	Автоматична ...	Модернізован...	2000	DarkSupply	2024-11-09	Продано
6	6	Меч	Холодна зброя	Традиційний ...	500	OldSwordMaster	2024-11-07	В наявності
7	7	Ремінгтон 870	Помповий ...	Міцний ...	700	WeaponMaster	2024-11-05	В наявності
8	8	ППШ-41	Штурмова ...	Історична ...	900	ColdWarGuns	2024-11-03	Продається

Рисунок 3.13 – Робота програми

Детальніше можна побачити елементи на рисунку 3.14, які були оброблені для зручнішого перегляду всередині програми, без необхідності лізти в excel.

	ID	Product Name	Weapon Type	Description	Price (USD)	Seller	Date Added	Status
1	1	AK-47	Автоматична ...	Військовий ...	1500	DarkDealer	2024-11-16	В наявності
2	2	Глок 19	Пістолет	Популярний ...	800	ShadowGuns	2024-11-15	В наявності
3	3	AR-15	Гвинтівка	Використовує...	1200	BlackMarketKing	2024-11-14	Продається
4	4	Снайперська ...	Снайперська ...	Далекобійна ...	5000	SilentSniper	2024-11-10	В наявності
5	5	M-16	Автоматична ...	Модернізован...	2000	DarkSupply	2024-11-09	Продано

Рисунок 3.14 - Виведення інформації

На рисунку вище видно, що було стягнуто елементи, проте невідомо з яких сайтів, все через те, що поля .onion не заповнювались. Виправлений результат приведено на рисунку 3.15.

	ID	Product Name	Weapon Type	Description	Price (USD)	Seller	Date Added	Status	URL
1	1	AK-47	Автоматична ...	Військовий ...	1500	DarkDealer	2024-11-16	В наявності	http://cash2tuohd4juo2l6us2ya64qbjtavwtmydb7txbraq7m6qs36dc5tqd.onion/
2	2	Глок 19	Пістолет	Популярний ...	800	ShadowGuns	2024-11-15	В наявності	http://...
3	3	AR-15	Гвинтівка	Використовує...	1200	BlackMarketKing	2024-11-14	Продається	http://...

Рисунок 3.15 – Покращення виведення інформації

Для покращення візуалу також було добавлено пошук, який дозволить шукати цікаві елементи вже поміж спаршених елементів. На рисунку 3.16 зобарежно результат пошуку по елементах завдяки пошуку, тобто фільтр для минулого результат парсера.

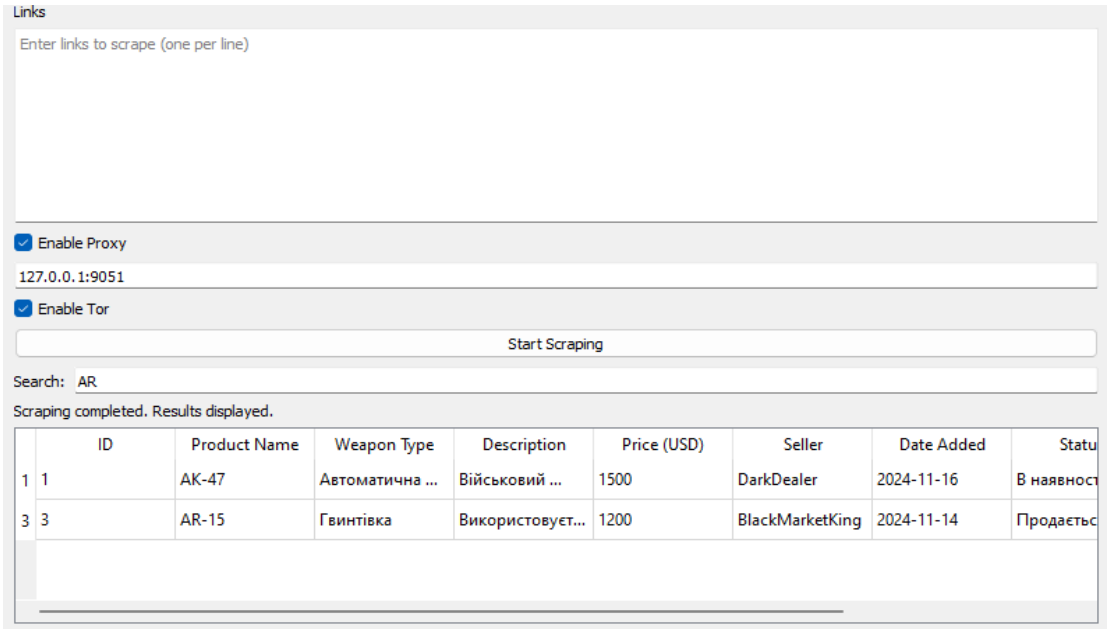
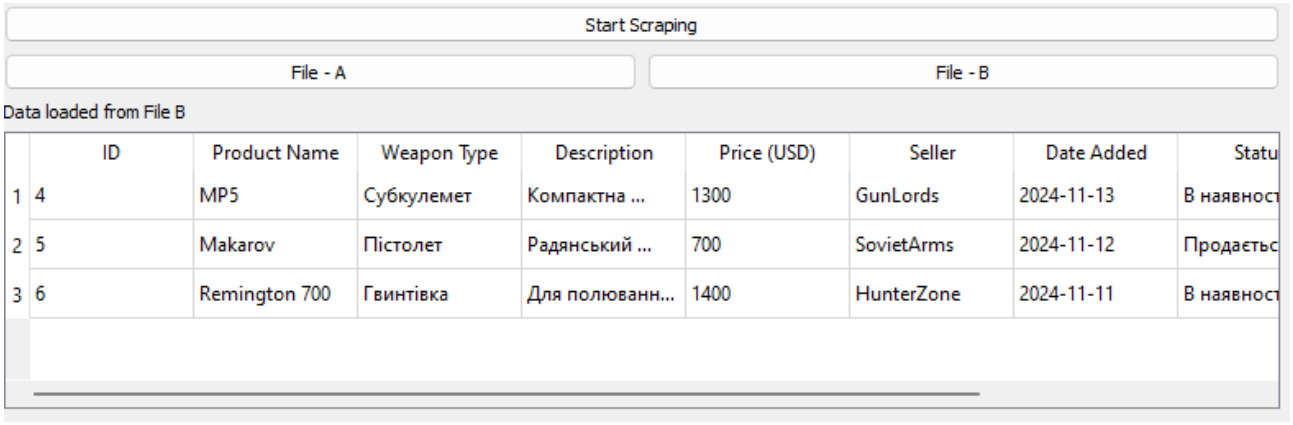


Рисунок 3.16 – Пошук по запиту

Після було добавлено ще один пункт, що зображено на рисунку 3.17.



	ID	Product Name	Weapon Type	Description	Price (USD)	Seller	Date Added	Statu
1	4	MP5	Субкулемет	Компактна ...	1300	GunLords	2024-11-13	В наявност
2	5	Makarov	Пістолет	Радянський ...	700	SovietArms	2024-11-12	Продаєтьс
3	6	Remington 700	Гвинтівка	Для полюванн...	1400	HunterZone	2024-11-11	В наявност

Рисунок 3.17 – Пошук по файлах

Це пошук по файлах, щоб було зрозуміло, які файли зараз обробляє парсер, адже файли з різних пагінацій можуть відрізнятись в секунди, або ж виникне в майбутньому потреба порівнювати зміни в паринґу файлах за тиждень, допустим, або відкрити файли парсингу за 11.16.13.52.15 і 11.16.14.52.32, для розуміння було використано файли a.csv і b.csv.

Висновки до розділу 3.

1. Проведено розробку програмного користувацького інтерфейсу для системи моніторингу дарквеб-ресурсів. Реалізовано інтуїтивно зрозумілий інтерфейс, який забезпечує зручне управління процесами збору та аналізу даних. Інтерфейс розроблено з використанням сучасних технологій та фреймворків, що дозволяє ефективно візуалізувати отримані дані та керувати налаштуваннями системи. Особлива увага приділена розробці компонентів для відображення статистики, графіків та звітів на основі зібраної інформації.

2. Створено систему моніторингу дарквеб-ресурсів з використанням технологій автоматизованого підключення до мережі Tor. Розроблено модулі для автоматичного встановлення з'єднання, зміни IP-адрес та управління налаштуваннями проксі. Реалізовано механізми обробки помилок та відновлення з'єднання при збоях, що забезпечує стабільну роботу системи. Інтегровано

бібліотеки для роботи з Tor Controller Protocol, що дозволяє гнучко керувати параметрами з'єднання.

3. Реалізовано компоненти для збору та аналізу даних з дарквеб-ресурсів. Розроблено парсери для обробки HTML-структур різної складності, механізми емуляції користувацької активності та системи подолання захисту від автоматизованого збору даних. Створено модулі для обробки динамічного контенту та асинхронного завантаження даних. Реалізовано алгоритми фільтрації та категоризації отриманої інформації для подальшого аналізу.

4. Розроблено механізми формування та обробки CSV-файлів для зберігання зібраних даних.

ВИСНОВКИ

1. Проведено теоретичний аналіз Dark Web та діпнету, розглянуто їх сутність, характеристики та основні компоненти, а також проаналізовано підходи інших експертів до визначення цих понять.

2. Визначено та проаналізовано основні джерела інформації, що потребують моніторингу для ефективного виявлення та запобігання кіберзлочинності у Dark Web.

3. Проведено детальний аналіз основних видів кіберзлочинності в Dark Web, розглянуто специфіку та різновиди послуг і товарів.

4. Проаналізовано правові аспекти України використання джерел для моніторингу Dark Web, також розглянуто закони інших держав.

5. Проведено аналіз технологій анонімного доступу та безпеки, які широко використовуються в Dark Web для забезпечення конфіденційності та захисту користувачів. Зокрема, детально розглянуто протокол Tor (The Onion Router), який є основою анонімної мережі Dark Web.

6. Проаналізовано технології автоматизованого підключення до мережі Tor, які дозволяють спростити процес налаштування та управління з'єднаннями з Dark Web'ом.

7. Розглянуто системи скрапінгу даних з Dark Web, які дозволяють автоматизувати процес збору інформації з прихованих веб-ресурсів. Проаналізовано основні підходи та інструменти для скрапінгу даних, такі як використання парсерів HTML/XML, емуляція дій користувача за допомогою headless-браузерів (наприклад, Puppeteer або Selenium), а також обробка динамічного контенту та подолання антискрапінгових механізмів. Досліджено особливості застосування цих технологій в контексті Dark Web, де додатково необхідно враховувати анонімність, безпеку та обхід специфічних обмежень.

8. Описано алгоритм пошуку даних з джерел, зібраних в Dark Web, на основі формату CSV (Comma-Separated Values). Запропонований алгоритм дозволяє ефективно обробляти та аналізувати великі обсяги структурованих даних, отриманих в результаті скрапінгу Dark Web.

9. Проведено розробку програмного користувацького інтерфейсу для системи моніторингу дарквеб-ресурсів. Реалізовано інтуїтивно зрозумілий інтерфейс, який забезпечує зручне управління процесами збору та аналізу даних.

10. Створено систему моніторингу дарквеб-ресурсів з використанням технологій автоматизованого підключення до мережі Tor. Розроблено модулі для автоматичного встановлення з'єднання, зміни IP-адрес та управління налаштуваннями проксі.

11. Створено модулі для обробки динамічного контенту та асинхронного завантаження даних. Реалізовано алгоритми фільтрації та категоризації отриманої інформації для подальшого аналізу.

12. Розроблено механізми формування та обробки CSV-файлів для зберігання зібраних даних. Реалізовано алгоритми очищення та нормалізації даних, видалення дублікатів та форматування інформації відповідно до визначеної структури.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Christin N. Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace // WWW '13: Proceedings of the 22nd international conference on World Wide Web. Rio de Janeiro: ACM Press, 2013. С. 213-224. DOI: 10.1145/2488388.2488408.

2. Moore D. Cryptopolitik and the Darknet [Електронний ресурс] / D. Moore, T. Rid. – 2016. – Режим доступу до ресурсу: https://www.tandfonline.com/doi/full/10.1080/00396338.2016.1142085#d1e106.

3. Biryukov A. Trawling for Tor Hidden Services: Detection, Measurement, Deanonymization [Електронний ресурс] / A. Biryukov, I. Pustogarov, R. Weinmann // IEEE. – 2013. – Режим доступу до ресурсу: https://ieeexplore.ieee.org/document/6547103.

4. Bartlett J. The Dark Net / Jamie Bartlett. – London: Random House, 2014. – 320 с.

5. Owen G. The Tor Dark Net. Global Commission on Internet Governance Paper Series / G. Owen, N. Savage., 2015. – 20 с.

6. Aldridge, J., & Décary-Hétu, D. Not an 'eBay for Drugs': The Cryptomarket 'Silk Road' as a Paradigm Shifting Criminal Innovation / [Електронний ресурс]. - Режим доступу: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2436643

7. Europol. Internet Organised Crime Threat Assessment (IOCTA) / [Електронний ресурс]. - Режим доступу: https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment

8. Ф. Боргоново, А. Фішер, Г. Порріно, С. Р. Лучіні та Л. Мартиньйон, «Виявлення цілей мережі прихованих послуг темної мережі за допомогою моніторингу та аналізу соціальних мереж», *2023 Міжнародний семінар IEEE з технологій для захисту та безпеки (TechDefense)* , Рим, Італія, 2023, стор. 412-416, doi:

9. Tor Metrics / [Електронний ресурс]. - Режим доступу: https://metrics.torproject.org/

10. RAND Corporation. Dark Web Intelligence Report / [Електронний ресурс]. - Режим доступу: https://www.rand.org/pubs/research_reports/RR2704.html

11. INTERPOL. Darknet and Cryptocurrencies / [Електронний ресурс]. - Режим доступу: https://www.interpol.int/en/Crimes/Cybercrime/Darknet-and-cryptocurrencies

12. Jardine E. The Dark Web Dilemma: Tor, Anonymity and Online Policing / E. Jardine. - Global Commission on Internet Governance Paper Series No. 21. - [Електронний ресурс]. - Режим доступу: https://www.cigionline.org/publications/dark-web-dilemma-tor-anonymity-and-online-policing/

13. Dingledine R., Mathewson N., Syverson P. Tor: The Second-Generation Onion Router / Proceedings of the 13th USENIX Security Symposium. - [Електронний ресурс]. - Режим доступу: https://dl.acm.org/doi/10.5555/1251375.1251396

14. Exploit.in: форум для обміну інформацією про вразливості / [Електронний ресурс]. - Режим доступу: https://exploit.in/

15. Mirea M., Wang V., Jung J. "The not so dark side of the darknet: a qualitative study" / Security Journal, 31. - [Електронний ресурс]. - Режим доступу: https://link.springer.com/article/10.1057/s41284-018-0150-5

16. Alnabulsi H., Islam R. Identification of Illegal Forum Activities Inside the Dark Net / International Conference on Machine Learning and Data Engineering (iCMLDE). - Sydney, NSW, Australia, 2018. - [Електронний ресурс]. - Режим доступу: https://ieeexplore.ieee.org/document/8613998

17. Mikhaylov A., Frank R. Dark Cards and Money: Analysis of Online Money Laundering Schemes / European Intelligence and Security Informatics Conference (EISIC). - Uppsala, Sweden, 2016. - [Електронний ресурс]. - Режим доступу: https://ieeexplore.ieee.org/document/7870194

18. Kruithof K., Aldridge J., Décary-Hétu D., Sim M., Dujso E., Hoorens S. "Internet-facilitated drugs trade: An analysis of the size, scope and the role of the Netherlands" / RAND Corporation, Santa Monica, CA. - [Електронний ресурс]. - Режим доступу: https://www.rand.org/pubs/research_reports/RR1607.html

19. Tor Project / [Електронний ресурс]. - Режим доступу: https://www.torproject.org/

20. The Hidden Wiki / [Електронний ресурс]. - Режим доступу: https://thehiddenwiki.org/

21. I2P: децентралізована анонімна мережа / [Електронний ресурс]. - Режим доступу: https://geti2p.net/en/

22. Freenet Project / [Електронний ресурс]. - Режим доступу: https://freenetproject.org/

23. Group-IB Threat Intelligence / [Електронний ресурс]. - Режим доступу: https://www.group-ib.com/intelligence/

24. Clarke I., Hong T.W., Miller S.G., Sandberg O., Wiley B. "Protecting Free Expression Online with Freenet" / IEEE Internet Computing, vol. 6, no. 1. - [Електронний ресурс]. - Режим доступу: https://ieeexplore.ieee.org/document/978368

25. FireEye Threat Intelligence Reports / [Електронний ресурс]. - Режим доступу: https://www.mandiant.com/resources/reports

26. Chainalysis Crypto Crime Report / [Електронний ресурс]. - Режим доступу: https://www.chainalysis.com/reports/

27. Закон України "Про основні засади забезпечення кібербезпеки України" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/2163-19

28. Закон України "Про захист інформації в інформаційно-телекомунікаційних системах" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/80/94-вр

29. Закон України "Про захист персональних даних" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/2297-17

30. Постанова КМУ "Про затвердження Порядку формування переліку об'єктів критичної інформаційної інфраструктури та Порядку внесення об'єктів критичної інформаційної інфраструктури до державного реєстру об'єктів критичної інформаційної інфраструктури" від 9 жовтня 2020 р. № 943 /

[Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/943-2020-п

31. ENISA Threat Landscape Report / [Електронний ресурс]. - Режим доступу: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023

32. FBI Internet Crime Report / [Електронний ресурс]. - Режим доступу: https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf

33. Cybersecurity & Infrastructure Security Agency / [Електронний ресурс]. - Режим доступу: https://www.cisa.gov/cybersecurity

34. Microsoft Digital Defense Report / [Електронний ресурс]. - Режим доступу: https://www.microsoft.com/en-us/security/business/security-intelligence-report

35. Google Threat Analysis Group / [Електронний ресурс]. - Режим доступу: https://blog.google/threat-analysis-group/

36. Kaspersky Lab Darknet Intelligence Report / [Електронний ресурс]. - Режим доступу: https://securelist.com/

37. Закон України "Про доступ до публічної інформації" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/2939-17

38. Recorded Future Dark Web Analysis / [Електронний ресурс]. - Режим доступу: https://www.recordedfuture.com/

39. Кримінальний кодекс України / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/2341-14

40. Закон України "Про інформацію" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/2657-12

41. Кримінальний кодекс України, стаття 361 "Несанкціоноване втручання в роботу електронно-обчислювальних машин (комп'ютерів), автоматизованих систем, комп'ютерних мереж чи мереж електрозв'язку" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/2341-14#n2508

42. Кримінальний кодекс України, стаття 361-1 "Створення з метою використання, розповсюдження або збуту шкідливих програмних чи технічних засобів, а також їх розповсюдження або збут" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/2341-14#n2510

43. Center for Internet Security / [Електронний ресурс]. - Режим доступу: https://www.cisecurity.org/insights/white-papers

44. Конвенція про кіберзлочинність / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/994_575

45. The Onion Router: Official Documentation / [Електронний ресурс]. - Режим доступу: https://2019.www.torproject.org/docs/documentation.html

46. McAfee Advanced Threat Research / [Електронний ресурс]. - Режим доступу: https://www.mcafee.com/enterprise/en-us/threat-center.html

47. Symantec Internet Security Threat Report / [Електронний ресурс]. - Режим доступу: https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence

48. Digital Shadows Dark Web Research / [Електронний ресурс]. - Режим доступу: https://www.digitalshadows.com/blog-and-research/

49. Check Point Research / [Електронний ресурс]. - Режим доступу: https://research.checkpoint.com/

50. Trend Micro Forward-Looking Threat Research / [Електронний ресурс]. - Режим доступу: https://www.trendmicro.com/en_us/research.html

51. Palo Alto Networks Unit 42 Threat Research / [Електронний ресурс]. - Режим доступу: https://unit42.paloaltonetworks.com/

52. Конвенція Ради Європи про кіберзлочинність (Convention on Cybercrime), Стаття 2-10 / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/994_575#n60

53. PyQt5 Documentation / [Електронний ресурс]. - Режим доступу: https://www.riverbankcomputing.com/static/Docs/PyQt5/

54. Конвенція Ради Європи про кіберзлочинність (Convention on Cybercrime), Стаття 15 "Умови і гарантії" / [Електронний ресурс]. - Режим доступу: https://zakon.rada.gov.ua/laws/show/994_575#n143

55. Python Documentation: Web Scraping with BeautifulSoup / [Електронний ресурс]. - Режим доступу: https://www.crummy.com/software/BeautifulSoup/bs4/doc/

56. CSV File Format Specification / [Електронний ресурс]. - Режим доступу: https://tools.ietf.org/html/rfc4180

57. О-О. РОМАНЮК. ПРАВОВІ ТА ЕТИЧНІ АСПЕКТИ МОНІТОРИНГУ ТЕМНОГО ВЕБУ. Збірник матеріалів науково-практичного симпозіуму «Захист інформації», Тернопіль, 2024. С. 48-50

58. О-О. РОМАНЮК. ІНСТРУМЕНТИ ТА МЕТОДИ МОНІТОРИНГУ ВІДКРИТИХ ДЖЕРЕЛ. Збірник матеріалів науково-практичної конференції молодих вчених, аспірантів та студентів «Кібербезпека та комп’ютерно-інтегровані технології» (КБКІТ - 2024), Тернопіль, 2024. С. 140-142

ДОДАТОК А

Код PyQT5

```
import sys
from PyQt5.QtWidgets import QApplication, QWidget, QLabel, QLineEdit, QTextEdit, QPushButton, QCheckBox, QVBoxLayout, QHBoxLayout, QMessageBox

class DarkwebScraperUI(QWidget):
    def __init__(self):
        super().__init__()
        self.setWindowTitle('Darkweb Scraper')
        self.setGeometry(100, 100, 400, 400)
        self.initUI()

    def initUI(self):
        layout = QVBoxLayout()

        fields_label = QLabel('Fields to Scrape')
        self.fields_input = QTextEdit()
        self.fields_input.setPlaceholderText('Enter fields to scrape (one per line)')

        keywords_label = QLabel('Keywords')
        self.keywords_input = QLineEdit()
        self.keywords_input.setPlaceholderText('Enter keywords (comma-separated)')

        links_label = QLabel('Links')
        self.links_input = QTextEdit()
        self.links_input.setPlaceholderText('Enter links to scrape (one per line)')

        self.proxy_checkbox = QCheckBox('Enable Proxy')
```

```
    self.proxy_checkbox.stateChanged.connect(self.toggle_proxy)
    self.proxy_input = QLineEdit()
    self.proxy_input.setPlaceholderText('Enter proxy URL')
    self.proxy_input.setVisible(False)

    self.tor_checkbox = QCheckBox('Enable Tor')

    scrape_button = QPushButton('Start Scraping')
    scrape_button.clicked.connect(self.start_scraping)

    self.status_label = QLabel()

    layout.addWidget(fields_label)
    layout.addWidget(self.fields_input)
    layout.addWidget(keywords_label)
    layout.addWidget(self.keywords_input)
    layout.addWidget(links_label)
    layout.addWidget(self.links_input)
    layout.addWidget(self.proxy_checkbox)
    layout.addWidget(self.proxy_input)
    layout.addWidget(self.tor_checkbox)
    layout.addWidget(scrape_button)
    layout.addWidget(self.status_label)

    self.setLayout(layout)

def toggle_proxy(self, state):
    if state == 2:
        self.proxy_input.setVisible(True)
    else:
        self.proxy_input.setVisible(False)
```

```
    def start_scraping(self):
        fields = self.fields_input.toPlainText().strip().split('\n')
        keywords = self.keywords_input.text().strip().split(',')
        links = self.links_input.toPlainText().strip().split('\n')
        proxy_enabled = self.proxy_checkbox.isChecked()
        proxy_url = self.proxy_input.text().strip() if proxy_enabled else None
        tor_enabled = self.tor_checkbox.isChecked()

        self.status_label.setText('Scraping in progress...')
        QApplication.processEvents()
        self.status_label.setText('Scraping completed. Results saved to CSV.')

if __name__ == '__main__':
    app = QApplication(sys.argv)
    scraper_ui = DarkwebScraperUI()
    scraper_ui.show()
    sys.exit(app.exec_())
```

ДОДАТОК Б

Копії публікацій



ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ПОЛІТЕХНІКА»
ГАЛИЦЬКИЙ ФАХОВИЙ КОЛЕДЖ ІМ. В'ЯЧЕСЛАВА ЧОРНОВОЛА

КІБЕРБЕЗПЕКА ТА КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ (КБКІТ – 2024)

науково-практична конференція
молодих вчених, аспірантів та студентів

26–28 серпня 2024
Тернопіль

Збірник матеріалів науково-практичної конференції молодих вчених, аспірантів та студентів «Кібербезпека та комп'ютерно-інтегровані технології» (КБКІТ - 2024), Тернопіль, 2024. - 160 с.

Адреса редакції:
Західноукраїнський національний університет, кафедра кібербезпеки,
вул. Олени Теліги 8, м. Тернопіль 46003
Контакти:
*e-mail:*conferencekb@gmail.com

УДК 004.056(477)

Орест-Остап РОМАНЮК

Західноукраїнський національний університет

ІНСТРУМЕНТИ ТА МЕТОДИ МОНІТОРИНГУ ВІДКРИТИХ ДЖЕРЕЛ

Вступ. Розвиток інформаційних технологій і поширення цифрових комунікацій вимагають ефективного моніторингу інформації з відкритих джерел. Вивчення інструментів та методів аналізу опублікованих даних стає все більш важливим у різних сферах, від кібербезпеки до журналістських розслідувань.

Мета: проаналізувати сучасні інструменти та методологічні підходи до моніторингу відкритих джерел, визначити їхню ефективність та особливості практичного застосування в різних галузях.

1. Основні інструменти та методи моніторингу відкритих джерел для виявлення загроз на темному вебу

Дослідження темного вебу та його потенційних загроз становить складну технічну задачу, яка вимагає глибоких знань у сфері кібербезпеки та розвідувальних технологій. Специфіка моніторингу передбачає використання спеціалізованого програмного забезпечення, яке дозволяє здійснювати анонімний доступ та аналіз захищених інформаційних середовищ. Мережа Tor, що є основною інфраструктурою темного вебу, створює додаткові технічні бар'єри для традиційних методів розвідки та моніторингу. Складність полягає в необхідності подолання multiplex-шифрування, яке унеможливлює миттєву ідентифікацію джерел та учасників комунікацій [1].

На рисунку 1 зображено схему взаємодії основних інструментів моніторингу темного вебу. Схема відображає послідовність і взаємозв'язки між компонентами, що забезпечують процес збору, аналізу та обробки інформації з прихованих мереж.

Рисунок 1 - Схема взаємодії інструментів моніторингу темного вебу

Ключовими інструментами для моніторингу темного вебу виступають спеціалізовані краулери та агрегатори інформації, розроблені з урахуванням специфіки анонімних мереж. Серед найбільш ефективних рішень можна виділити програмні комплекси Maltego, Memex, та SpiderFoot, які дозволяють здійснювати комплексний збір та кореляцію розрізнених інформаційних слідів. Технічна архітектура цих інструментів передбачає використання складних алгоритмів машинного навчання для автоматичної класифікації потенційно небезпечного контенту. Принципово важливим є застосування техніки fingerprinting - унікальної ідентифікації інформаційних об'єктів без розкриття первинного джерела [2].

Методологія розвідки темного вебу включає декілька послідовних етапів технічної розвідки. Первинний збір даних здійснюється через спеціалізовані браузери з підтримкою анонімізації, такі як Tor Browser або Tails. Наступним кроком виступає технічна кореляція зібраних інформаційних слідів з використанням штучного інтелекту, здатного розпізнавати приховані патерни

комунікацій. Машинне навчання дозволяє класифікувати потенційні загрози за рівнем ризику, враховуючи контекст, лінгвістичні особливості та мережеву поведінку учасників [3].

На рисунку 2 зображено архітектуру розподіленої системи моніторингу, що складається з таких ключових компонентів:

Рисунок 2 - Архітектура розподіленої системи моніторингу

Технічні обмеження традиційних інструментів розвідки спонукають дослідників до розроблення гібридних підходів, що поєднують технології комп'ютерної лінгвістики, нейронних мереж та статистичного аналізу. Принципово новим напрямком є використання квантових алгоритмів для декодування складних криптографічних послідовностей темного вебу, що дозволяє здійснювати більш глибокий аналіз латентних комунікативних каналів. Важливим елементом моніторингу темного вебу є верифікація отриманих даних, що передбачає використання методів перевірки достовірності інформації. Це може включати аналіз метаданих, перевірку джерел та вивчення контексту, в якому інформація була отримана. Застосування технологій блокчейн для фіксації даних про транзакції та комунікації може суттєво підвищити рівень довіри до зібраної інформації. Водночас, важливо враховувати етичні аспекти моніторингу, оскільки анонімність користувачів темного вебу може бути порушена, що викликає питання про конфіденційність та права людини [4].

Окрім того, для виявлення загроз на темному вебу використовуються методи соціальної інженерії, які дозволяють дослідникам отримувати інформацію шляхом взаємодії з учасниками анонімних мереж]. Це може включати створення фальшивих профілів або участь у закритих форумах, де обговорюються незаконні дії. Такі методи вимагають високого рівня обережності та знань про специфіку комунікацій у темному вебі, оскільки будь-яка помилка може призвести до викриття особи дослідника.

На рисунку 3 зображено графік залежності типів загроз, який демонструє взаємозв'язки між основними видами незаконної діяльності, що здійснюється у темному вебі. Важливим є також використання аналітичних платформ, які дозволяють візуалізувати дані та виявляти зв'язки між різними об'єктами. Це може включати графічні інтерфейси, які демонструють мережі взаємодії між учасниками, а також інструменти для аналізу тексту, що дозволяють виявляти ключові слова та фрази, пов'язані з загрозами. Візуалізація даних допомагає не лише виявляти загрози, але й формувати стратегії реагування на них.

Рисунок 3 - Графік залежності типів загроз

Висновок. Вивчення інструментів і методів моніторингу темного Інтернету показує надзвичайну складність і багатогранність процесу виявлення потенційних загроз в анонімному мережевому середовищі. Технологічний ландшафт сучасної кібербезпеки вимагає постійного вдосконалення підходів до розвідки та аналізу потоків інформації, що циркулюють за межами традиційного інтернет-простору. Еволюція інструментів моніторингу безпосередньо пов'язана з розвитком штучного інтелекту, машинного навчання та квантових обчислювальних технологій. Практичне застосування розглянутих технологій показує, що ефективність моніторингу темного інтернету залежить від комплексного підходу, який поєднує в собі технічні інструменти, алгоритми машинного аналізу і глибоке розуміння деталей середовища анонімного спілкування. Важливою особливістю сучасних розвідувальних технологій є здатність не тільки фіксувати інформаційні сліди, а й прогнозувати потенційний напрямок розвитку загроз. Технічні обмеження, пов'язані з багаторівневим шифруванням та анонімізацією комунікацій, спонукають дослідників постійно вдосконалювати свій методологічний підхід. Інтегруючи різні інструменти, від спеціалізованих сканерів до нейронних мереж, ви можете сформувати більш повне уявлення про обробку інформації в темній мережі. У той же час етичні та юридичні аспекти такого стеження вимагають особливої уваги і дотримання балансу між необхідністю забезпечення безпеки і повагою конфіденційності користувачів.

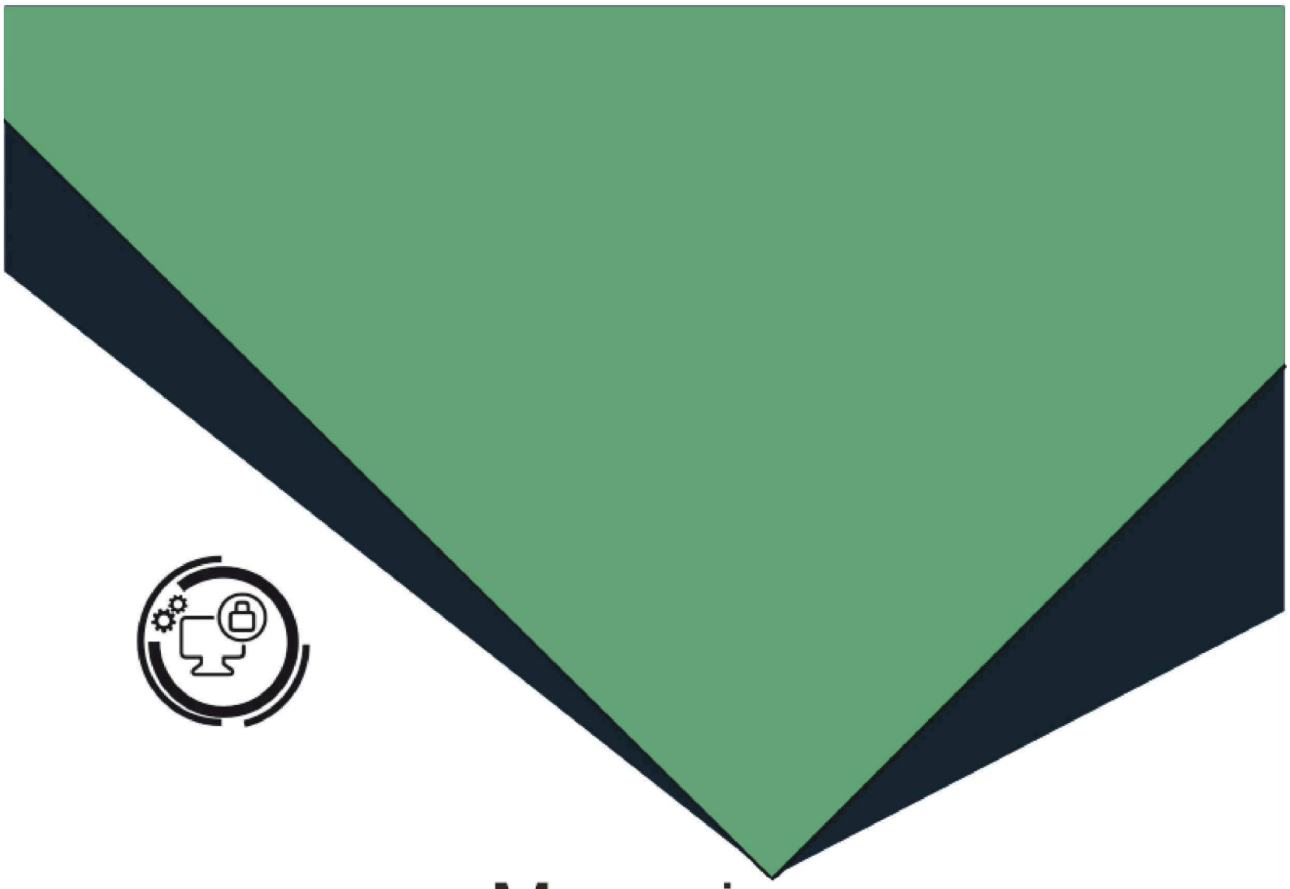
Перелік використаних джерел.

1. Tiwari, Shiva Verma, Ravi Jaiswal, Janvi Rai, Bipin Kumar. (2020). Open Source Intelligence Initiating Efficient Investigation and Reliable Web Searching. 151-163. 10.1007/978-981-15-6634-9_15.

2. What is Cyber Threat Intelligence? [Електронний ресурс].- Режим доступу: https://www.crowdstrike.com/en-us/cybersecurity-101/threat-intelligence/?srsltid=AfmBOopOlO5IslR2vDLs_tGLzNMfb2wag82U2IMhQfajoZbRPgbTy6pP

3. Brenner, S. W. (2013). Cybercrime: Criminal Threats from Cyberspace. Routledge. [Електронний ресурс].- Режим доступу: https://ecommons.udayton.edu/cgi/viewcontent.cgi?article=1023context=law_fac_pub

4. Mäntymäki, M., Riemer, K. (2016). Enterprise social networking: A knowledge management perspective. International Journal of Information Management, 36(6, Part A), 1042–1052.



Матеріали науково-практичного симпозіуму «ЗАХИСТ ІНФОРМАЦІЇ»

2024

ГРОМАДСЬКА ОРГАНІЗАЦІЯ
«КІБЕРБЕЗПЕКА І АВТОМАТИЗАЦІЯ»

Матеріали
науково-практичного симпозіуму
«ЗАХИСТ ІНФОРМАЦІЇ»

30 листопада 2024
Тернопіль

У збірнику опубліковано матеріали науково-практичного симпозіуму «Захист інформації», Тернопіль, 2024. - 130с.

Технічний редактор: Давлетова А.Я.

Адреса редакції:
Громадська організація «Кібербезпека і автоматизація»
м. Тернопіль
Контактний телефон: (066)043-42-10
e-mail: conferencekb@gmail.com

ЗМІСТ

УДК 004.056(477)

Орест-Остап РОМАНЮК

Західноукраїнський національний університет

ПРАВОВІ ТА ЕТИЧНІ АСПЕКТИ МОНІТОРИНГУ ТЕМНОГО ВЕБУ

Вступ. Темний веб (darknet) є частиною інтернету, яка використовує шифрування та анонімні мережі для приховування діяльності своїх користувачів. Хоча темний веб надає важливі можливості для забезпечення конфіденційності та свободи слова, він також часто асоціюється з незаконною діяльністю, такою як торгівля наркотиками, зброєю або навіть порнографією за участю неповнолітніх. У зв'язку з цим, питання моніторингу та регулювання темного вебу стає все більш актуальним і викликає гострі дебати щодо балансу між безпекою, приватністю та свободою.

Мета: аналіз правових та етичних аспектів моніторингу темного вебу, зокрема правові рамки, які регулюють моніторинг темного вебу, включаючи питання конфіденційності, свободи слова та кібербезпеки.

1. Правові межі моніторингу темного вебу

Темний веб кидає виклик традиційним правовим рамкам, оскільки його анонімність та зашифрованість ускладнюють правове регулювання. З одного боку, право на конфіденційність та свободу слова гарантують громадянам можливість використовувати темний веб для захисту своєї приватності та висловлення поглядів. З іншого боку, анонімність темного вебу створює сприятливе середовище для незаконної діяльності, такої як торгівля наркотиками, зброєю або навіть сексуальна експлуатація дітей. Це викликає занепокоєння щодо безпеки та законності і змушує уряди шукати шляхи моніторингу та регулювання темного вебу [1].

Правові рамки, що регулюють моніторинг темного вебу, є предметом активних дискусій та суперечок. Одним із ключових питань є баланс між захистом конфіденційності та забезпеченням безпеки. Згідно з Загальним регламентом ЄС про захист даних (GDPR), громадяни мають право на захист своїх персональних даних, включаючи анонімність в інтернеті. Однак, відповідно до законодавства про національну безпеку, правоохоронні органи можуть отримувати доступ до певної інформації в темному вебі для розслідування злочинів. Знайти правильний баланс між цими двома пріоритетами є серйозним викликом для законодавців [2].

Ще одним важливим аспектом є свобода слова. Темний веб часто використовується як платформа для висловлення неортодоксальних або навіть суперечливих поглядів, які можуть бути заборонені в інших місцях. На рисунку 1 зображено баланс між конфіденційністю. Питання полягає в тому, чи має уряд право обмежувати доступ до такого контенту, навіть якщо він вважається неприйнятним або небезпечним. Прихильники свободи слова стверджують, що будь-які спроби цензури або моніторингу темного вебу є загрозою демократії та основним правам людини [3].

Крім того, правові обмеження на моніторинг темного вебу можуть мати наслідки для кібербезпеки. Хоча анонімність темного вебу може використовуватися зловмисниками, вона

також надає важливі можливості для захисту вразливих груп, таких як правозахисники або журналісти-розслідувачі, від онлайн-загроз. Сліпе застосування моніторингу може завдати шкоди законним користувачам темного вебу та наразити їх на небезпеку [4].

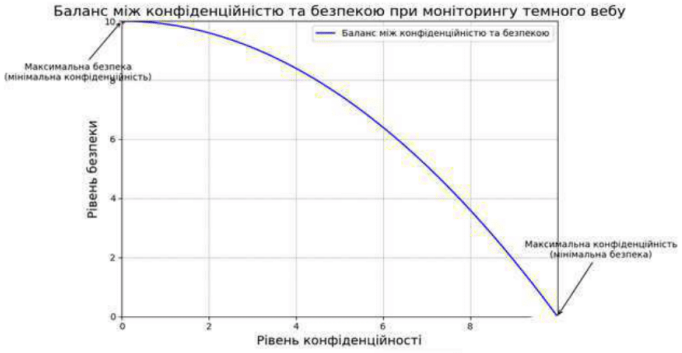


Рисунок 1 - Баланс між конфіденційністю та безпекою при моніторингу темного вебу

Пошук балансу між захистом конфіденційності, свободою слова та кібербезпекою є складним завданням, що стоїть перед законодавцями. Для вирішення цих суперечливих питань необхідний ретельний аналіз та широке обговорення з усіма зацікавленими сторонами, включаючи правоохоронні органи, постачальників інтернет-послуг, експертів з прав людини та громадянське суспільство. Лише за допомогою всеохоплюючого підходу можна знайти ефективні правові рамки, які б поважали основні права та свободи, забезпечуючи при цьому безпеку суспільства.

У контексті темного вебу також важливо звертати увагу на міжнародний аспект, оскільки інтернет не має кордонів, а це ускладнює правове регулювання з боку окремих держав. Наприклад, контент, який є законним у одній країні, може бути забороненим в іншій. Це створює ускладнення у взаємодії між країнами у боротьбі з кіберзлочинністю та регулюванні контенту. Міжнародні угоди та співпраця між правоохоронними органами різних країн є необхідними для ефективного моніторингу та розслідування незаконної діяльності в темному вебі.

Додатково, технології моніторингу продовжують розвиватися, що відкриває нові можливості для виявлення злочинної діяльності. Наприклад, нові алгоритми аналізу великих даних і штучного інтелекту можуть виявляти тенденції та аномалії у поведінці користувачів, що дозволяє правоохоронним органам та організаціям безпеки своєчасно реагувати на потенційні загрози. Проте ці технології також піднімають етичні питання щодо приватності та можливості зловживання владою.

Однак важливо, щоб правові рамки, які визначають використання таких технологій, були максимально прозорими та підзвітними. Залучення громадян до обговорення законодавства про моніторинг темного вебу може допомогти створити більш балансований підхід та уникнути зловживань. Важливо також розробити механізми контролю, які б гарантували дотримання норм безпеки і прав людини, уникаючи дискримінаційного або невиправданого втручання в приватне життя громадян.

Зрештою, боротьба з незаконною діяльністю на темному вебі не повинна відбуватися

за рахунок фундаментальних прав та свобод. Знайти правильний баланс між захистом суспільства, захистом конфіденційності та свободою слова є викликом, який вимагатиме від законодавців гнучкості, врахування нових технологій та зворотного зв'язку від суспільства. Лише шляхом конструктивного діалогу та співпраці можна досягти вирішення цієї складної пробеми та створити правову базу, яка буде відповідати викликам сучасності й забезпечить безпеку всіх учасників інформаційного простору.

Висновок. Моніторинг і регулювання «темного інтернету» є складним і суперечливим питанням, яке вимагає дотримання ретельного балансу між захистом основних прав і свобод громадян та забезпеченням громадської безпеки. З одного боку, анонімність і шифрування Темного Інтернету надають важливі можливості для захисту приватності, свободи вираження поглядів і кібербезпеки, особливо для вразливих груп, таких як правозахисники і журналісти-розслідувачі. Однак, з іншого боку, така анонімність створює сприятливе середовище для незаконної діяльності та загрожує громадській безпеці.

Законодавча база, що регулює спостереження в темному інтернеті, перебуває в постійній напрузі між захистом приватного життя, свободою слова та кібербезпекою. Перед законодавцями стоїть завдання сформулювати чіткі правила, які забезпечують достатні гарантії для захисту основних прав громадян, і в той же час визначають умови, за яких правоохоронні органи можуть отримати доступ до інформації в темному інтернеті. Важливо знайти баланс між забезпеченням адекватного рівня безпеки та недопущенням надмірної цензури і порушень прав людини.

Вирішення таких суперечливих питань вимагає інклюзивного та комплексного підходу із залученням широкого кола зацікавлених сторін, включаючи правоохоронні органи, інтернет-провайдерів, експертів з прав людини та громадянське суспільство. Лише через постійний діалог, компроміс та адаптацію до мінливих реалій можна створити ефективну правову базу, яка б поважала основні права і свободи, забезпечуючи при цьому соціальну безпеку. Процес має бути прозорим і підзвітним, щоб громадяни могли контролювати дії влади та оскаржувати будь-які порушення. Загалом, вирішення проблеми моніторингу темного Інтернету вимагає творчого мислення, гнучкості та готовності постійно переглядати підходи. Лише за таких умов можна знайти баланс між захистом фундаментальних прав людини та безпекою в цій складній сфері. Це важливе питання з далекосяжними наслідками для майбутнього інтернету і нашого суспільства в цілому.

Перелік використаних джерел.

1. Owen G., Savage N. The Tor Dark Net // Global Commission on Internet Governance Paper Series. - 2015. - №20. [Електронний ресурс].- Режим доступу: https://www.cigionline.org.

2. Horton-Eddison M., Di Cristofaro M. Understanding Darknet Markets: Opportunities for HCI Research // Proceedings of the ACM on Human-Computer Interaction. - 2020. - Vol. 4. - doi:10.1145/3432936.

3. Soska K., Christin N. Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem // Proceedings of the 24th USENIX Security Symposium. - 2015. [Електронний ресурс].- Режим доступу: https://www.usenix.org.