УДК 343.85
МРНТИ 10.81.71

А.А. Калиев
Академия правоохранительных органов при Генеральной прокуратуре Республики Казахстан, г. Косшы, Республика Казахстан

АНАЛИЗ ОТКРЫТЫХ ДАННЫХ В ЭПОХУ ЦИФРОВЫХ ТЕХНОЛОГИЙ: ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ OSINT

Аннотация. Современные технологии неизбежно привели к увеличению объемов данных, размещаемых в цифровом пространстве. Это, в свою очередь, сделало разведку на основе открытых источников (OSINT) одним из важных и востребованных инструментов интернет-аналитики. Автор раскрыл историю становления OSINT, его эволюцию от первых попыток использования открытых источников до современных методов автоматизированного сбора и анализа данных. В данной работе, помимо современных методов и способов получения и обработки данных с помощью OSINT, автором рассматриваются конкретные аналитические инструменты поиска и анализа информации: Google Dorking, Maltego и Shodan. Также в статье особое внимание уделяется вопросу использования искусственного интеллекта и программ машинного обучения в развитии направления OSINT. Автор приходит к выводу о том, что дальнейшее развитие OSINT будет определяться не только технологическим прогрессом, но и необходимостью регулирования данного процесса для предотвращения злоупотреблений и защиты персональных данных.

Ключевые слова: OSINT; открытый источник; интернет; база данных; искусственный интеллект; машинное обучение; поисковая система; метаданные.

А.Ә. Қалиев
Қазақстан Республикасы Бас прокуратурасы жанындағы Құқық қорғау органдары академиясы, Қосшы қ., Қазақстан Республикасы

ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР ДӘУІРІНДЕГІ АШЫҚ ДЕРЕКТЕРДІ ТАЛДАУ: OSINT МҮМКІНДІКТЕРІ МЕН ПЕРСПЕКТИВАЛАРЫ

Аннотация. Заманауи технологиялар цифрлық кеңістікте орналастырылатын деректер көлемінің ұлғаюына алып келді, бұл ашық көздерден барлау (OSINT) жүргізуді интернет-аналитиканың маңызды әрі сұранысқа ие құралдарының біріне айналдырды.

Автор OSINT қалыптасу тарихын ашып көрсетіп, ашық көздерді пайдаланудың алғашқы әрекеттерінен бастап, деректерді автоматтандырылған түрде жинау және талдау әдістеріне дейінгі эволюциясын сипаттайды. Сонымен қатар, осы еңбекте заманауи әдістер мен OSINT арқылы деректерді алу және өңдеу тәсілдерімен қатар, ақпаратты іздеу мен талдауға арналған нақты аналитикалық құралдар қарастырылады: Google Dorking, Maltego және Shodan.

Сондай-ақ, мақалада жасанды интеллект пен машиналық оқыту бағдарламаларын OSINT дамуы аясында қолдану мәселесіне ерекше назар аударылады. Автор OSINT-тің болашақтағы дамуы тек технологиялық прогреске ғана емес, сонымен қатар осы процесті реттеу қажеттілігіне де байланысты болатынын атап өтеді. Бұл реттеу шаралары ықтимал теріс пайдалану жағдайларының алдын алу және жеке деректерді қорғау мақсатында маңызды рөл атқарады.

Түйінді сөздер: OSINT; ашық көз; интернет; дерекқорлар; жасанды интеллект; машиналық оқыту; іздеу жүйесі; метадеректер.

A.A. Kaliyev
The Law Enforcement Academy under the Prosecutor General's Office of the Republic of Kazakhstan, Kosshy, the Republic of Kazakhstan

ANALYSIS OF OPEN DATA IN THE DIGITAL AGE: OPPORTUNITIES AND PROSPECTS OF OSINT

Abstract. Modern technologies have inevitably led to an increase in the volume of data placed in the digital space, making open-source intelligence (OSINT) one of the most important and in-demand tools in internet analytics. The author explores the history of OSINT, tracing its evolution from the first attempts to utilize open sources to modern methods of automated data collection and analysis. In addition to contemporary methods of obtaining and processing data through OSINT, the study examines specific analytical tools for information search and analysis: Google Dorking, Maltego, and Shodan. Special attention is also given to the role of artificial intelligence and machine learning programs in the development of OSINT. The author concludes that the future of OSINT will be determined not only by technological progress but also by the need for regulatory measures to prevent misuse and ensure the protection of personal data.

Keywords: OSINT; open source; internet; database; artificial intelligence; machine learning; search engine; metadata.

DOI: 10.52425/25187252_2025_37_98

Введение. В современном мире, окутанном всевозможными цифровыми устройствами умение находить нужную информацию является одним из самых важных навыков, помогающих принимать обоснованные решения, расследовать онлайн и офлайн преступления, анализировать и сопоставлять различные факты, выявляя скрытые связи между людьми и произошедшими событиями. При этом источниками информации могут служить не только печатные издания, открытые базы данных, но и интернет-ресурсы, включая социальные сети, форумы, блоги, а также данные, полученные из метаданных файлов, изображений и видео.

Благодаря стремительному развитию и внедрению в повседневную жизнь человека современных технологий, а также распространению цифровых платформ, объем доступной для анализа информации каждый день увеличивается в геометрической прогрессии, что делает открытую разведку Open Source Intelligence (далее – OSINT) мощным аналитическим инструментом, в т.ч. при проведении расследований.

Люди, не задумываясь, выкладывают в интернет огромное количество информации о себе, включая фото и видеоматериалы, ежедневно оставляя цифровые следы в социальных сетях, на форумах, в базах данных, размещенных на специальных сайтах. Помимо этого, значительный объем данных генерируют государственные органы и многочисленные организации.

Однако эффективный сбор и обработка данных требуют от специалиста или сотрудника правоохранительного органа не только знаний о доступных инструментах, но и определенных навыков работы с ними, умения верифицировать информацию и правильно ее применять.

Цель статьи – рассмотреть основные методы и инструменты OSINT, применяемые для поиска, анализа и интерпретации данных, примеры их использования, а также выявить возможности и перспективы использования разведки на основе открытых источников.

Материалы и методы. Для исследования темы научной статьи использованы методы системного анализа, описания, интерпретации и обобщения практики использования инструментов OSINT, в т.ч. непосредственно автором статьи. Теоретической основой послужили научные публикации отечественных и зарубежных авторов. В качестве эмпирического материала использовались реальные кейсы из журналистских расследований, а также данные об использовании таких инструментов, как Google Dorking, Maltego и сервисы обратного поиска изображений.

Результаты, обсуждение. Открытые источники – это все то, что можно свободно найти и прочитать или посмотреть без каких-либо специальных разрешений.

В различных источниках по-разному определяется понятие OSINT. Одни авторы рассматривают этот термин исключительно как метод сбора информации из общедоступных источников, другие рассматривают это понятие более шире, включая анализ, интерпретацию и использование полученных данных для принятия решений.

Так, например, М. Баззелл в книге «Open Source Intelligence» определяет OSINT как процесс поиска, сбора и анализа данных из открытых источников для последующего использования [1]. В свою очередь С. Бертрам в своей работе «The Tao of Open Source Intelligence» подчеркивает, что OSINT как открытая разведка – это не просто сбор информации, а разведывательная деятельность, включающая в себя помимо вышеуказанного поиска, сбора и анализа информации, еще структурирование и оценку полученных данных [2].

Тем самым, OSINT – это многогранное понятие, которое в большей степени зависит от контекста его применения.

Если обращаться к истории OSINT, то она далеко уходит корнями в военную разведку, где сбор информации из открытых источников использовался еще за долго до цифровой эры. Первые упоминания об этом можно найти в деятельности армий и разведывательных служб, которые изучали иностранные печатные издания, радиопередачи, официальные заявления и дипломатические переписки для получения стратегически важной информации, включающей в себя, помимо разведданных, общественные настроения и военные маневры противника. Например, британская правительственная школа кодов и шифров анализировала информацию из перехваченных сообщений Советского Союза, чтобы выявить стратегические намерения другого государства[1].

Однако настоящая революция в области разведки по открытым данным произошла уже с появлением глобальной сети интернет и новых технологий.

С учетом достижений науки и техники современные информационные системы сейчас позволяют уже в режиме реального времени отслеживать события, анализировать цифровые следы преступников и даже прогнозировать угрозы на основе анализа больших данных. Тем самым, OSINT эволюционировал от сбора сведений из газет и радио к сложным аналитическим системам, использующим искусственный интеллект и машинное обучение для работы с огромными массивами данных.

Сегодня OSINT, превратившись в самостоятельное направление, стал применяться не только государственными структурами, частными компаниями, военными и правоохранительными органами, но также и отдельными энтузиастами, журналистами.

В частности, современные журналисты используют методы разведки по открытым источникам для разоблачения фейковых новостей, расследования коррупционных схем и преступлений.

Например, одна из международных исследовательских групп «Bellingat» с помощью OSINT успешно раскрыла ряд резонансных дел, связанных с международными конфликтами, преступлениями и коррупцией. На официальном сайте «Bellingat» в разделе, посвященном MH17 собраны материалы и отчеты, подробно описывающие ход расследования и методы, использованные для анализа данных, что демонстрирует эффективность инструментов OSINT в современных журналистских расследованиях[2].

Также методы OSINT находят широкое применение и в правоохранительной деятельности, особенно в борьбе с организованной преступностью, терроризмом и киберпреступностью.

Правоохранительные органы разных стран все чаще обращаются к разведке по

[1] Спецслужбы на войне | Великобритания. — Все о Второй мировой [Электронный ресурс] – Режим доступа: https://wwii.space/%D0%A1%D0%BF%D0%B5%D1%86%D1%81%D0%BB%D1%83%D0%B6%D0%B1%D1%8B-%D0%BD%D0%B0-%D0%B2%D0%BE%D0%B9%D0%BD%D0%B5-%D0%92%D0%B5%D0%BB%D0%B8%D0%BA%D0%BE%D0%B1%D1%80%D0%B8%D1%82%D0%B0%D0%BD%D0%B8%D1%8F/ (дата обращения: 04.08.2025).

[2] MH17 и дезинформационные операции ГРУ, часть 1: медиапроект «Bonanza» – Беллингкэт [Электронный ресурс] – Режим доступа: https://ru.bellingcat.com/novosti/russia/2020/11/20/bonanza-gru/ (дата обращения: 04.08.2025).

открытым источникам, собирая информацию о подозреваемых, для выявления преступных схем и связей.

На сегодняшний день эффективность расследования любого преступления зависит от собранной по нему информации. Чем больше собрано оперативной информации, тем эффективнее будет идти расследование.

Американский аналитик Кен Шерман в одной из своих публикаций сказал о том, что правоохранительные и специальные службы США, да и многих других стран, большую часть оперативной информации, а это примерно около 80%, собирают из открытых источников, т.е. из «открытого» интернета, а оставшуюся часть – от 10 до 20%, черпают из ведомственных баз данных и от конфиденциальных помощников (агентов)[3].

Следовательно, одним из ключевых преимуществ OSINT для правоохранительных структур является его способность обеспечивать быстрый доступ к актуальной информации без необходимости получения санкций суда, что напоминает работу оперативных подразделений при проведении отдельных оперативно-розыскных мероприятий.

Рассматривая интернет в качестве площадки для получения информации, необходимо отметить его многоуровневость, которая включает в себя поверхностный (Surface Web), глубокий (Deep Web) и теневой (Dark Web) уровни, каждый из которых содержит различный объем данных и доступен с разной степенью сложности.

На рисунке № 1 наглядно показана структура интернета, включая его основные сегменты.

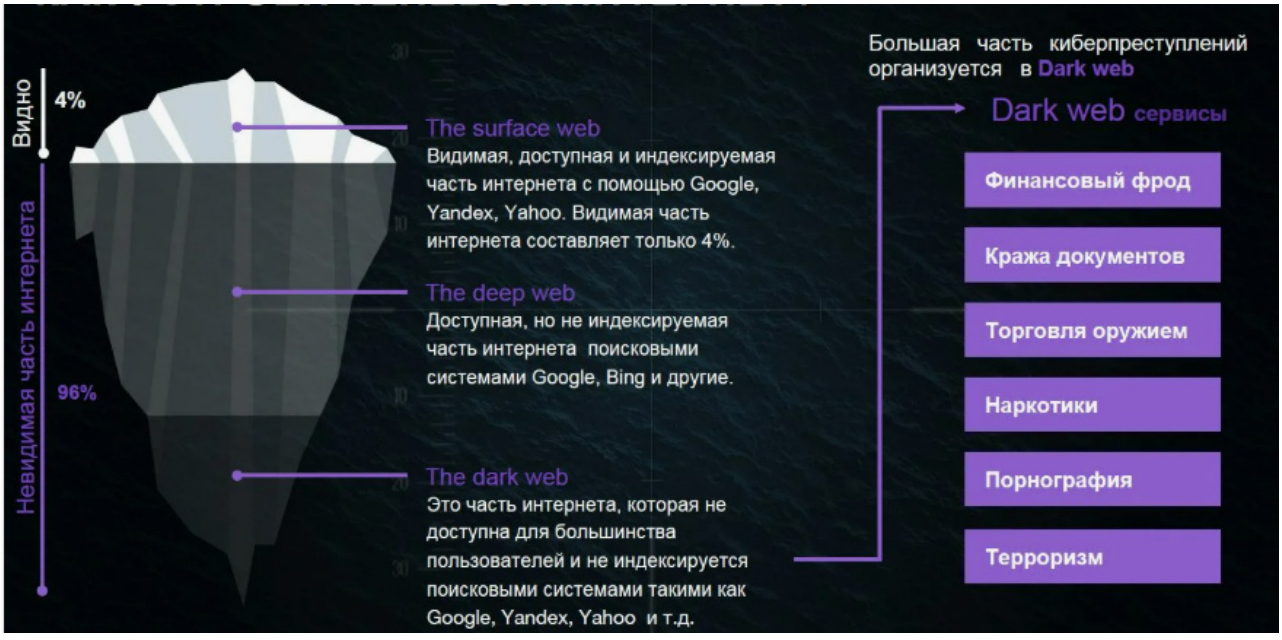

Рисунок № 1. Структура интернета

Поверхностный уровень – это открытая часть интернета, которая индексируется поисковыми системами, глубокий уровень представляет собой скрытую часть интернета и теневой уровень – это анонимизированная часть интернета, подключение к которой требует использование специального программного обеспечения.

Можно предположить, что теневой уровень интернета из-за своего названия используется лишь лицами, совершающими уголовные правонарушения или любые действия противоправного характера. Но на самом деле это не так. Теневой и глубокий уровни интернета, помимо вышеуказанных лиц, широко используют журналисты,

[3] Пробив: как определить близкое окружение // ВКонтакте [Электронный ресурс] – Режим доступа: https://vk.com/wall-131010772_3741?ysclid=m7wxwxj6e0945482711 (дата обращения: 04.08.2025).

исследователи, простые граждане для сохранения своей анонимности.

У каждого уровня есть свой объем информации, который содержится в базах данных интернета. Соответственно интернет – это хорошо структурированная система сбора, обработки и передачи информации, представляющая собой множество различных баз данных (поисковые системы, интернет-каталоги, блоги, форумы, социальные сети и т.д.).

Ищут и собирают информацию в интернете по-разному. Некоторые используют целый арсенал ресурсов всемирной паутины, другие ограничиваются лишь одним поисковиком. Нет общих стандартов поиска, определенных правил, как нужно искать информацию правильно. Но многие интернет-разведчики или интернет-аналитики придерживаются пяти основных шагов: 1) формирование задачи; 2) планирование; 3) сбор информации; 4) анализ информации; 5) формирование выводов.

Работа с OSINT по сбору информации должна начинаться с четко сформулированной задачи. Например, анализ деятельности компании, проверка информации о человеке, мониторинг общественно-политической и социально-экономической ситуаций.

На данном этапе нужно четко понимать цель поиска: какая информация нужна и как ее можно использовать.

Следующий шаг предусматривает планирование, где уже специалист определяет направления поиска и источники информации. Примерные направления для поиска представлены на рисунке № 2.

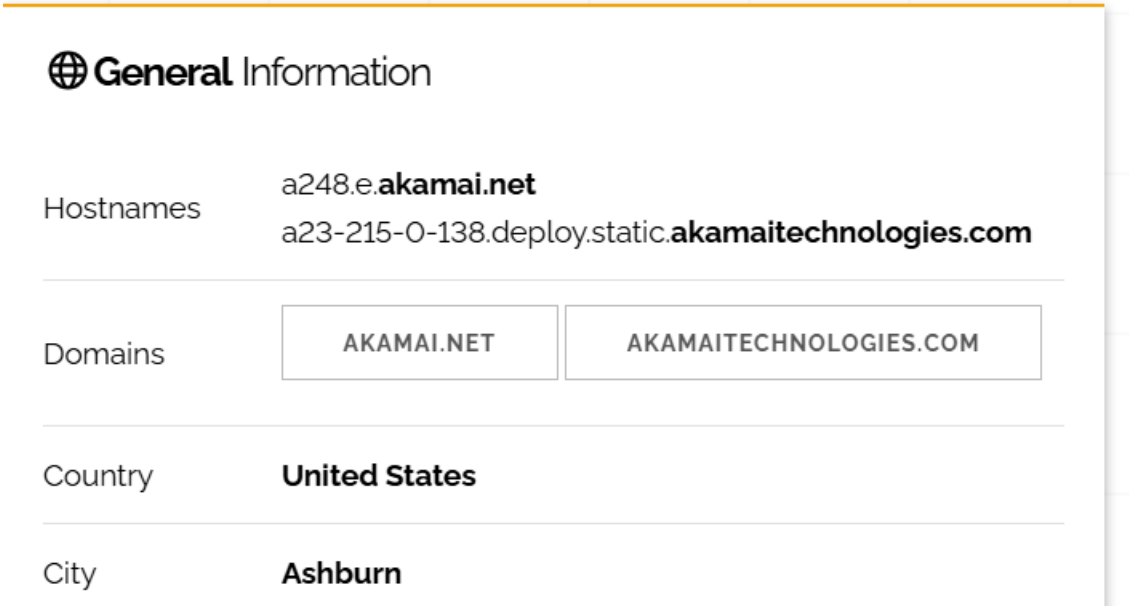

Рисунок № 2. Направления поиска информации

Как видно, направления для поиска информации очень разнообразные: это социальные сети, поисковые системы, форумы, специальные сайты, интернет-каталоги, сайты государственных органов и частных организаций.

Поисковые системы помогают находить скрытые страницы и документы, социальные сети раскрывают связи между людьми, сайты государственных органов и частных организаций раскрывают информацию о них.

Следующий шаг – это поиск и сбор информации. Поиск информации – это процесс извлечения хранимой информации, удовлетворяющей информационные потребности, т.е. являющейся релевантной [3]. Релевантная информация – это та информация, которая необходима для решения некоторой конкретной задачи [4].

Например, о субъекте поиска можно собрать следующую информацию: ФИО, число, месяц, год рождения, место рождения, место проживания, ИИН, круг его связей, интересы, увлечения, номера резервных мобильных устройств.

После этого можно переходить к анализу

результатов. Анализировать нужно фотографии, переписки, публичные сообщения или записи, круг друзей, интересы, место жительства, учебы или работы и многое др.

И завершающим шагом должна стать подготовка выводов. В результате поиска собирается информация, по которой можно как минимум получить психологический портрет человека. Это анкетные данные, образование, место работы, род деятельности, кто его родители и в каких социальных сетях он представлен, а также где он работает и чем занимается, сколько он зарабатывает, какие у него есть активы и другие данные, необходимые для начала или продолжения расследования.

Однако сбор информации – это лишь начальный процесс, за которым следует другой – проверка достоверности, поскольку иногда данные бывают недостоверными или устаревшими. В этом случае эксперты по интернет-разведке используют метод кросс-верификации. Данный термин активно используется в таких сферах, как кибербезопасность, OSINT, анализ данных и означает процесс проверки информации, данных или определенных результатов с использованием нескольких независимых источников для подтверждения достоверности.

То есть, они сравнивают информацию из нескольких независимых источников, анализируют временные метки публикаций, проводят обратный поиск изображений и т.д.

Например, если найденная фотография имеет отметку о том, что сделана в 2023 году, но ее оригинал находится в сети интернет с 2018 года, это повод для аналитиков задуматься о возможной манипуляции с данными. Еще один важный аспект проверки достоверности является анализ метаданных файлов, который раскрывает скрытую изначально информацию о дате съемки, местоположении и даже устройстве, на котором был создан документ.

Помимо ручного поиска широко используются специализированные программы, такие как Google Dorking, Shodan, Maltego, I2 и др. Данные инструменты позволяют автоматизировать процесс сбора информации, ускоряя процесс поиска данных, анализировать сетевые ресурсы и устанавливать взаимосвязи между различными объектами расследования.

Например, Maltego и I2 (ай ту) строят графовые модели связей между людьми, организациями и доменами в удобном виде, позволяя анализировать сложные схемы взаимодействия. На рисунке № 3 изображена схема такого взаимодействия.

Рисунок № 3. Схема взаимодействия между субъектами (организациями)

То есть современные технологии предоставляют широкий спектр инструментов, автоматизирующих процесс сбора информации.

В рамках данной статьи можно рассмотреть конкретный пример использования наиболее востребованных инструментов поиска и сбора информации. Допустим, следователь обнаружил в социальной сети фотографию человека, по которой необходимо собрать данные. Чтобы определить, где еще может встречаться данное изображение, используются сервисы обратного поиска: «Google Reverse Image Search», «Yandex Images» или «TinEye».

Фотография загружается в один из вышеуказанных сервисов, который выдает результат в виде перечня интернет-ресурсов. Также предположим, что в ходе проверки одного из них установлена ссылка на неизвестный следствию сайт игровых автоматов, например, www.example.com.

Зная доменное имя сайта, следователь может попытаться определить его IP-адрес с помощью командной строки. Для этого запускается командная строка путем одновременного нажатия клавиш «Windows» и «R» и введения команды «cmd». После этого в командной строке необходимо ввести слово: «ping example.com», которая выведет на экран IP-адрес данного сайта – 23.215.0.138.

Далее, используя один из сервисов получения информации по IP-адресу, например «Shodan», следователь получает общую информацию: город, страна, организация, интернет-провайдер, как показано на рисунке № 4 ниже.

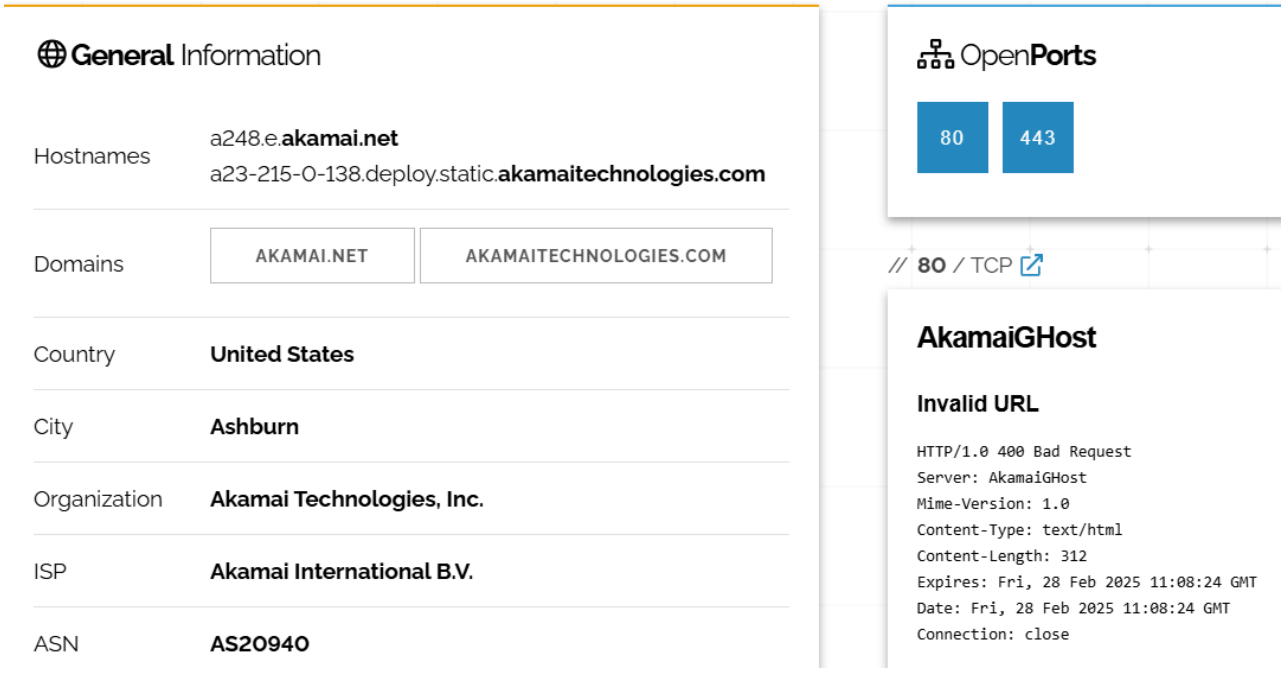

Рисунок № 4. Информация с программы «Shodan»

Если сервис показал географическое положение IP-адреса можно использовать карты и спутниковые снимки для анализа района, зданий и возможных объектов, относящихся к искомому лицу.

Тем самым данный процесс демонстрирует, как с помощью методов OSINT можно собрать цифровой след, оказывающий помощь в расследовании.

Заключение. Обобщая изложенное, можно отметить, что разведка на основе OSINT является незаменимым инструментом современного расследования и аналитической деятельности, а также занимает все более важное место в современных условиях цифровой среды. Она уже давно вышла за рамки обычного вспомогательного метода и превратилась

в полноценный аналитический инструмент, активно применяемый как в государственном, так и в частном секторе. Сложность и многоуровневость интернета требуют от специалистов высокой квалификации, технической грамотности и развитого аналитического мышления.

В этой связи предлагается создать в структурах правоохранительных органов подразделения по OSINT-разведке, нацеленные на поддержку и сопровождение расследований. Также рекомендуется разработать и внедрить в образовательный процесс ВУЗов правоохранительной направленности учебные модули, направленные на формирование у будущих следователей и оперативных сотрудников практических навыков поиска, сбора и анализа информации, в т.ч. с использованием специализированного программного обеспечения.

Список использованной литературы:

1. Bazzell, M. Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information; 7th edition. / M. Bazzell. – St. Louis: CreateSpace Independent Publishing Platform, 2018. – 622 p.
2. Bertram, S. The Tao of Open Source Intelligence / S. Bertram. – Лондон: IT Governance Publishing, 2015. – 200 p.
3. Макарова, Н.В. Информатика: учебник для вузов / Н.В. Макарова, В.Б. Волков. – СПб.: Питер, 2011. – 576 с.
4. Шокин, Ю.И. Проблемы поиска информации / Ю.И. Шокин, А.М. Федотов, В.Б. Барахнин. – Новосибирск: Наука, 2010. – 220 с.

References:

1. Bazzell, M. Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information; 7th edition. / M. Bazzell. – St. Louis: CreateSpace Independent Publishing Platform, 2018. – 622 p.
2. Bertram, S. The Tao of Open Source Intelligence / S. Bertram. – London: IT Governance Publishing, 2015. – 200 p.
3. Makarova, N.V. Informatika: uchebnik dlja vuzov / N.V. Makarova, V.B. Volkov. – SPb.: Piter, 2011. – 576 s.
4. Shokin, Ju.I. Problemy poiska informacii / Ju.I. Shokin, A.M. Fedotov, V.B. Barahnin. – Novosibirsk: Nauka, 2010. – 220 s.

АВТОРЛАР ТУРАЛЫ МӘЛІМЕТТЕР / СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Асқар Әбужанұлы Қалиев – Қазақстан Республикасы Бас прокуратура жанындағы Құқық қорғау органдары академиясының Кәсіптік оқыту институтының жаһандық қатерлерге қарсы іс-қимыл жөніндегі арнайы даярлық кафедрасының доценті, e-mail: askar909@mail.ru.

Калиев Аскар Абужанович – доцент кафедры специальной подготовки по противодействию глобальным угрозам Института профессионального обучения Академии правоохранительных органов при Генеральной прокуратуре, e-mail: askar909@mail.ru.

Kaliyev Askar Abuzhanovich – Associate Professor of the Department of Special Training in Countering Global Threats at the Institute of Professional Training of the Law Enforcement Academy under the Prosecutor General's Office of the Republic of Kazakhstan, e-mail: askar909@mail.ru.