УДК 004.8:004.738.5

АВТОМАТИЗАЦИЯ ПРОЦЕССОВ OSINT: РАЗРАБОТКА КОМПЛЕКСНЫХ РЕШЕНИЙ ДЛЯ СБОРА И АНАЛИЗА ОТКРЫТЫХ ДАННЫХ

Дубников В.А. (ВКА), Ратушняк И.А. **(ВКА)**, Тельбух В.В. **(ВКА)**
Научный руководитель – кандидат технических наук, преподаватель Тельбух В. В.
(Военно-космическая академия имени А.Ф.Можайского)

Введение. Современный мир характеризуется ростом объёмов открытых данных, важных для кибербезопасности, маркетинга, журналистики и управления репутацией. Ручной сбор информации неэффективен, что подчеркивает необходимость автоматизации OSINT (Open Source Intelligence).

OSINT — это сбор и анализ данных из общедоступных источников: социальных сетей, веб-сайтов и публичных реестров. Автоматизация OSINT ускоряет сбор, повышает точность, минимизирует ошибки и обеспечивает глубокий анализ с использованием современных технологий [1].

Цель работы — разработка комплексного решения для автоматизации OSINT, включающего сбор, обработку, анализ и визуализацию данных. Предложены методы для создания универсальной платформы, адаптируемой к различным типам данных и задачам.

Основная часть. Ручной сбор и анализ данных из открытых источников сопряжен с существенными ограничениями, включая высокую трудоемкость, низкую скорость обработки и субъективность интерпретации результатов. Существующие инструменты OSINT, несмотря на их функциональность, зачастую не способны эффективно обрабатывать большие объемы данных, интегрировать информацию из разнородных источников и обеспечивать глубокий аналитический анализ. Эти ограничения актуализируют необходимость разработки усовершенствованных решений, основанных на автоматизации и применении современных технологий [2,3].

В рамках исследования предложен комплексный подход к автоматизации процессов OSINT, структурированный в виде интерактивной карты инструментов, сгруппированных по категориям. Пользователь сможет выбрать определенный класс, перейти к подклассу и выбрать конкретный источник информации. Такой подход позволит систематизировать доступ к инструментам и значительно упростить процесс сбора и анализа данных.

Комплекс будет включать четыре ключевых этапа:

1. Сбор данных. Автоматизация извлечения информации из открытых источников (социальные сети, новостные порталы, базы данных) с использованием технологий веб-скрейпинга и API. Это позволяет охватить большие объемы информации и минимизировать ручной труд.

Примеры:

- Shodan: Поисковая система для обнаружения устройств, подключенных к интернету, включая серверы, камеры и IoT-устройства, по IP-адресам, портам или доменам;
- SpiderFoot: Автоматизированный инструмент для сбора данных из более чем 100 источников, включая социальные сети, WHOIS, DNS и другие.

2. Обработка данных. Очистка и предварительная подготовка данных, включая удаление дубликатов, фильтрацию нерелевантной информации и нормализацию текста. Такая обработка обеспечивает высокое качество данных для последующего анализа.

Примеры:

- OpenRefine: Инструмент для очистки и преобразования неструктурированных данных, таких как CSV или JSON;
- BeautifulSoup (Python): Библиотека для парсинга HTML/XML и извлечения нужной информации.

3. Анализ данных. Применение методов машинного обучения и обработки естественного языка (NLP) для классификации данных, выделения ключевых трендов и извлечения сущностей (имена, локации, организации).

Примеры:

- spaCy: Библиотека для NLP, которая позволяет извлекать сущности (имена, локации, организации) и анализировать текстовые данные;

- Maltego: Инструмент для визуализации и анализа связей между различными объектами, такими как люди, домены и IP-адреса.

4. Визуализация данных. Создание интерактивных отчетов и дашбордов для наглядного представления результатов анализа. Визуализация делает данные более доступными и понятными для пользователей.

Примеры:

- Tableau: Платформа для создания интерактивных дашбордов и визуализации данных;

- Gephi: Инструмент для визуализации графов связей между объектами, такими как люди, организации и события;

- Plotly: Библиотека для создания интерактивных графиков и диаграмм, которая поддерживает интеграцию с Python.

Экспериментальная проверка предложенного подхода на реальных данных (тексты из социальных сетей и новостных порталов) продемонстрировала его эффективность. Автоматизация процессов OSINT позволила значительно сократить временные затраты на обработку информации и повысить точность аналитических выводов, что подтверждает практическую применимость разработанного решения.

Выводы. Разработанное решение для автоматизации процессов OSINT демонстрирует высокую эффективность в сборе и анализе открытых данных за счет применения методов машинного обучения, обработки естественного языка (NLP) и технологий Big Data. Автоматизация обеспечивает значительное сокращение временных затрат и повышение качества аналитической обработки, что делает решение востребованным в различных областях.

В будущем планируется расширить функционал платформы, добавив сетевой анализ, прогнозирование и адаптацию под специфические задачи, что повысит её универсальность и эффективность.

Список использованных источников:

1. SkillFactory. Что такое OSINT: основные понятия и применение [Электронный ресурс]. – https://blog.skillfactory.ru/glossary/osint/ (дата обращения: 12.01.2025)

2. SpectrumData. OSINT: как работает разведка по открытым источникам [Электронный ресурс]. – URL: https://spectrumdata.ru/blog/proverka-soiskatelya/osint-kak-rabotaet-razvedka-po-otkrytym-istochnikam/ (дата обращения: 12.01.2025).

3. SecurityLab. OSINT: инструменты и методы разведки по открытым источникам [Электронный ресурс]. – https://www.securitylab.ru/blog/personal/Technolady/354845.php?ysclid=m6jkd2lhtq736157128 (дата обращения: 12.01.2025).