ЛОКАЛЬНО-АДАПТИВНЫЙ TELEGRAM-АГЕНТ ДЛЯ OSINT-РАЗВЕДКИ

Войткус И.А., студент гр.361401

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Белоусова Е.С. – канд. техн. наук, доцент

Аннотация. В статье рассматривается разработка локально-адаптивного Telegram-агента для автоматизации разведки по открытым источникам. Ввиду неэффективности западных платформ в локальном сегменте и ограничений решений, предложена архитектура на базе FastAPI и Telegram API. Инструмент использует локальную базу данных и включает модули лингвистической адаптации, генерации Dork-запросов и биометрической идентификации лиц DeepFace. Разработанная система обеспечивает информационный суверенитет, соответствует законодательству Республики Беларусь о защите персональных данных и позволяет безопасно, быстро формировать цифровое досье субъекта без передачи данных в зарубежные облачные сервисы.

Ключевые слова. кибербразведка, OSINT, Telegram-бот, биометрическая идентификация, FastAPI, защита персональных данных.

Разработка и применение систем автоматизированного сбора данных в Республике Беларусь строго регламентируются национальным законодательством. Согласно Закону РБ от 07.05.2021 № 99-З «О защите персональных данных», обработка общедоступной информации без согласия лица допускается (ст. 6), однако автоматизированная агрегация таких данных в единый цифровой профиль требует соблюдения принципа соразмерности и целесообразности.

Особое внимание при проектировании подобных систем уделяется работе биометрического модуля: согласно ст. 8 Закона № 99-З, обработка биометрических персональных данных (распознавание лиц посредством нейросетей) требует обязательного письменного согласия субъекта. Бесконтрольное использование разработанного агента для скрытой деанонимизации третьих лиц является неправомерным и может быть квалифицировано по ст. 203-1 УК РБ «Незаконные действия в отношении информации о частной жизни и персональных данных».

В условиях постоянного роста объемов информации в открытых источниках, важным является аудит информационной безопасности организации. Согласно методологии MITRIE ATT&CK, этап сбора информации по открытым источникам включает в себя тактику T1593 [1]. В предыдущем исследовании была продемонстрирована успешная апробация способа автоматического сбора данных с использованием платформы n8n [2]. Но в ходе анализа были выявлены ограничения: недостаточная гибкость логики обработки битых данных, зависимость от внешних триггеров запуска, а также проблем управления таймаутами. Было принято решение использовать Telegram бот для передачи входных данных. Основные причины данного выбора являются:

1. В отличии от обычного веб-сайта, где сервер имеет публичный IP-адрес, бот общается с серверами Telegram, что делает соединение невидимым для прямых атак.
2. Весь трафик между оператором и ботом защищен протоколами шифрования.
3. Бот обеспечивает мобильность не только пользователей персональных компьютеров, но и предоставляет возможность разведки со смартфона.

Встраивать такой бот в ноду созданной цепочки n8n было нецелесообразно, что объясняется тем, что при гибкой настройке n8n часто обрывает соединение, если процесс длится долго. В то время как в коде данную настройку можно гибко изменять.

Центральным элементом системы является хранилище данных university_db.json, использование формата JSON позволило обеспечить мгновенный парсинг данных средствами стандартных библиотек Python. Как и в предыдущем исследовании роль эталонного набора данных для тестирования локальной разведки, сыграли открытые данные о преподавателях с сайта iis.bsuir.by. Каждая запись включает в себя: идентификационные данные, метаданные, верификационные атрибуты. Такой подход обеспечил то, что данные не покидают периметр операционной системы, на которой все установлено. Блок-схема процесса взаимодействия с агентом представлена на рисунке 1.

Практическое применение популярных коммерческих и открытых платформ кибербразведки (такие как Maltego, SpiderFoot) приносит мало полезной информации на выходе в контексте локального сегмента сети в странах СНГ и Республике Беларусь. Данные таких систем в основном ориентированы на зарубежные публичные реестры и западные социальные сети. Другим фактором является лингвистический барьер, так как западные анализаторы некорректно обрабатывают кириллические метаданные, что приводит к обрыву цепочки поиска.

Рисунок 1 – Процесс взаимодействия с агентом

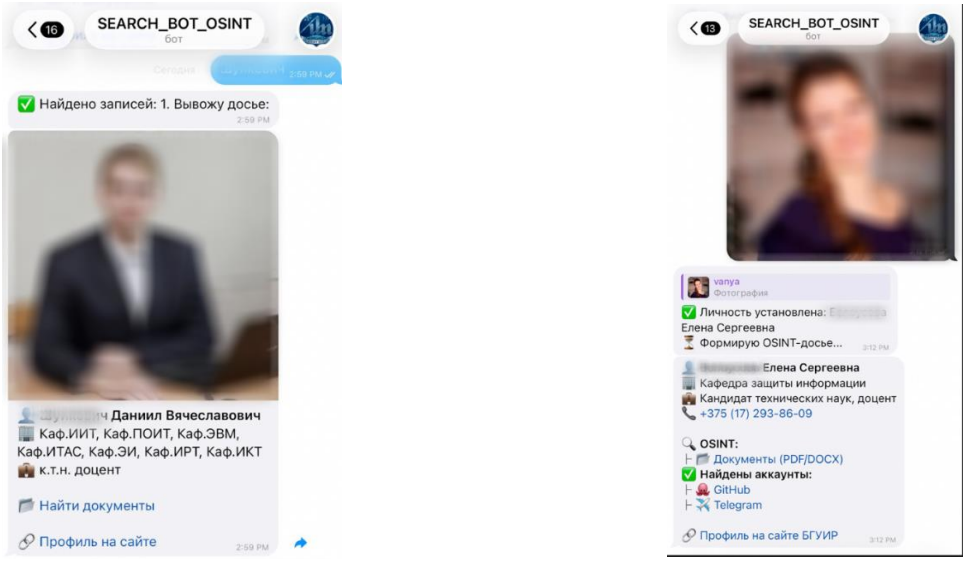
В связи с перечисленными факторами было принято решение об отказе использования готовых OSINT-платформ в пользу разработки собственного локально-адаптивного агента. Для этого были разработаны следующие модули:

1. Вычислительное ядро, в качестве основы которого выбран асинхронный веб-фреймворк FastAPI, его применение позволило реализовать нужную архитектуру и обеспечить параллельную обработку длительных сетевых таймаутов, возникающие при работе тяжеловесных модулей.

2. Клиентский интерфейс, представленный в виде Telegram бота, связь с которым представлена через Telegram API.

3. Модули лингвистической адаптации, использующие стандартные библиотеки Python (requests для HTTP-запросов и urllib для кодирования URL). А для предсказания возможных никнеймов был написан собственный модуль, генерирующий массивы никнеймов, которые могут подойти для дальнейшего поиска [3].

4. Модуль биометрической идентификации, реализованный с применением глубокого обучения DeepFace. В качестве вычислительной модели выбрана нейросеть Facenet для обеспечения вычисления векторных представлений лиц при работе с базой данных, созданной в начале.



а б

Рисунок 2 – Результат работы агента (а) и модуля идентификации по лицу (б)

Тестирование разработанного локально-адаптивного агента показало достаточно эффективные результаты. При вводе фамилии, имени и отчества или просто фамилии с, бот выдавал карточку преподавателя с информацией о должности, кафедре, номере телефона. Сгенерированные возможные никнеймы проверялись на наличие существующих аккаунтов в GitHub и Telegram, а также проводились специализированные Dork-запросы. Процесс занимает в среднем 1,5-2 секунды. Результаты проверки вышеописанных модулей, представленные ниже на рисунке 2, а.

Работа модуля биометрической идентификации реализована в соответствии со следующим алгоритмом:

1. Детектирование лица на загруженной фотографии с отсечением фона и программным выравниванием его по линии глаз для стандартизации ракурса.

2. Передача подготовленного изображения в нейросеть Facenet, которая преобразует уникальные черты лица в математический вектор, состоящий из 512 чисел.

3. Вычисление математического расстояния между полученным вектором и эталонными векторами, полученных из фотографий сотрудников БГУИР из локальной базы данных.

4. Анализ значения математического расстояния, если оно оказывается меньше заданного порога совпадения, алгоритм подтверждает личность и бот мгновенно формирует карточку досье.

Результат проверки такого компонента представлен на рисунке 3.

Таким образом, разработанный локально-адаптивный агент представляет собой не только эффективный инструмент OSINT-профилирования, но и стратегически важное решение для обеспечения информационного суверенитета, так как он учитывает специфику регионального цифрового следа (кириллица, локальные мессенджеры) и исключает риски трансграничной передачи данных в зарубежные облачные сервисы.

Список использованных источников:

1. Блэр, Р. Согласование операций безопасности с фреймворком MITRE ATT&CK / Р. Блэр. – Бирмингем: Packt Publishing, 2022. – 268 с. – ISBN 978-1804616697.

2. Войткус И.А. Инструмент для OSINT-разведки и реализации техники MITRIE ATTA&CK T1593 / И.А. Войткус, Е.С. Белоусова К.Ц. // Технические средства защиты информации : матер. XXIV Междунар. науч.-техн. конф. (Республика Беларусь, Минск, 08 апреля 2026 года) / редкол. : О. В. Бойправ [и др.]. – Минск : БГУИР, 2026. – С. 342–345.