

Пекишев Даниил Владимирович

**ОПТИМИЗАЦИЯ РАБОТЫ ПОИСКОВОГО СЕРВИСА В
TELEGRAM С ПРИМЕНЕНИЕМ СОВРЕМЕННЫХ МЕТОДОВ,
МОДЕЛЕЙ И АЛГОРИТМОВ ОБРАБОТКИ ИНФОРМАЦИИ**

В данной статье дается характеристика работы поискового сервиса в telegram, который функционирует на основе применения современных методов, моделей и алгоритмов обработки информации. Цель проекта – создание удобного сервиса по поиску учетных записей, оказавшихся в открытом доступе. Задача – разработка программного комплекса для успешного внедрения его в приложение, которое предоставляет информацию о компрометации личных данных пользователей в сети, что снижает риск взлома аккаунтов. Актуальность исследования заключается в том, что проблема компрометации личных данных в интернете огромна и ее решение требует разработки новых, инновационных программных продуктов. И этот сервис один из них.

Telegram-бот, инновационный сервис, алгоритмы обработки информации, методы обработки информации, потеря персональных данных, защита аккаунтов, поисковый бот, базы данных, оптимизация, анализ утечек.

Pekishev Daniil Vladimirovich

**OPTIMIZATION OF SEARCH SERVICE OPERATION IN TELEGRAM
USING MODERN METHODS, MODELS AND INFORMATION
PROCESSING ALGORITHMS**

This article describes the operation of the search service in telegram, which operates based on the use of modern methods, models and algorithms for information processing. The goal of the project is to create a convenient service for searching for accounts that are in the public domain. The task is to develop a software package for its successful implementation in an application that provides information about the compromise of users' personal data on the network, which reduces the risk of account hacking. The relevance of the study lies in the fact that the problem of compromising

personal data on the Internet is enormous and its solution requires the development of new, innovative software products. And this service is one of them.

Telegram bot, innovative service, information processing algorithms, information processing methods, loss of personal data, account protection, search bot, databases, optimization, leak analysis.

Введение

Telegram – это тот мессенджер, который без преувеличения считается одним из самых популярных и востребованных в мире. В наше время telegram-боты стали неотъемлемой частью нашей повседневной жизни, предлагая нам различные услуги, начиная от развлечений и заканчивая предоставлением ответов на серьезные поисковые запросы.

И чтобы обеспечить своевременную и точную обработку запрашиваемых данных, разработчикам ботов необходимо применять различные методы, модели и алгоритмы обработки информации, тем самым оптимизируя работу приложения, делая сервис более удобным при использовании, улучшая его качество и увеличивая скорость поиска [1,2,3,4,5].

Такие боты мгновенно обрабатывают запросы клиентов и намного сокращают время поиска информации для пользователя, что делает эти сервисы необходимыми в наш век высоких технологий и малого количества свободного времени у большинства людей [6]. Помимо этого, эти приложения специально разрабатываются для людей разных поколений: подростков, людей среднего и старшего возраста. Таким образом, аудитория клиентов огромна.

Кроме того, многие боты обрабатывают огромный массив личной информации пользователей, которая не должна оказаться в открытом доступе, поэтому к подобным программным продуктам предъявляют серьезные требования безопасности.

Мессенджеру telegram в этом отношении можно доверять, так как он характеризуется хорошей степенью защиты, надежен и прост при использовании.

Основная часть

Итак, в процессе работы над проектом был разработан telegram-бот по поиску учетных записей, оказавшихся в открытом доступе в интернете. Личные данные людей попадают в открытые источники по многим причинам – это и ошибки самих пользователей при работе в сети, и взломы киберпреступниками аккаунтов или сайтов с данными клиентов, также очень

часто причиной утечек данных служат вирусы и стиллеры на компьютерах пользователей, которые попадают на них вместе с программами.

Эти вирусы собирают сохранённые пароли из браузеров и другие данные, а потом отсылают их злоумышленникам.

Позже огромное количество таких баз собираются в коллекции «логов» и распространяются по интернету. В самых больших коллекциях сотни миллионов записей. При этом часто пользователи могут даже не догадываться, что их данные попали в открытый доступ таким образом, так как при обнаружении на компьютере вируса обычно его удаляют без смены паролей от аккаунтов на всех сохранённых сайтах. В дальнейшем полностью нерешенная проблема может привести к взломам этих аккаунтов и дополнительной компрометации личных данных.

В начале работы над проектом была поставлена конкретная задача: быстрое предоставление посредством telegram-бота полной информации о раскритиченных личных данных пользователя в сети. Для решения этой задачи была проделана огромная и кропотливая работа по определению всего функционала, по оформлению архитектуры сервиса, были написаны программы для создания визуального интерфейса и формирования удобного отчёта, а также разработаны алгоритмы быстрого поиска и преобразования найденных результатов в отчёт.

В дополнение к вышесказанному проведён анализ десятков тысяч уже попавших в открытый доступ утечек, из них собрана основная информация и проиндексирована для быстрого поиска. Проведена интеграция с другими сервисами, которая позволяет выполнять геолокацию по IP, дешифрирование паролей и визуализацию координат из утечек на карте.

Также при помощи сторонних сервисов осуществляется перевод интерфейса на различные языки. В результате работы была собрана база данных из утечек размером более 6 ТБ. Для того, чтобы можно было выполнить достаточно быстрый поиск по такому большому объёму данных, использованы алгоритмы оптимизации.

Для ускорения множественных запросов применяется кэширование результатов поиска и параллельная обработка. Кроме того, используются различные методы оптимизации доступа к файлам, например их отображение в память по mmap, а также кэширование уже отображённых файлов.

При нечётком поиске по множеству параметров используются эвристические алгоритмы генерации списка подзапросов и анализа их реле-

вантности, а также системы одновременного выполнения нескольких связанных запросов, если они имеют общие параметры.

Для анализа полученных данных применяются методы парсинга сырых данных и алгоритмы преобразования их сперва в удобный формат json, а затем в понятный людям отчёт [7]. При этом данные обогащаются информацией о названии и дате утечки, а также дополняются описанием обстоятельств этой утечки, что позволяет проще воспринять информацию из отчёта.

Для взаимодействия между программой и telegram-ботом используется алгоритм long-polling, который подразумевает регулярный опрос серверов telegram и получение списка запросов, отправленных в бот. В ответ на эти запросы сервис отправляет сформированные отчёты, причём время ответа занимает доли секунды. В качестве математических моделей планируется использование алгоритмов гибридных нейронных сетей с целью реализации нечёткого поиска по растровым изображениям и хэш-отображениям лиц.

Выводы

Таким образом, оптимизация работы поискового сервиса в telegram с использованием современных методов, моделей и алгоритмов обработки информации позволяет значительно улучшить качество поиска, сделать его более точным и быстрым, что, помимо прочего, гарантирует удобство при использовании приложения и подтверждает хороший функционал программного продукта [8,9,10].

Разработанный бот может по запросу (по ФИО, по почте, по телефону и по другим данным) показать, какие конфиденциальные данные конкретного человека раскрыты в сети, что позволит ему вовремя принять меры, защищающие его аккаунт от несанкционированного проникновения и использования этих данных ему во вред.

Бот даёт ёмкие и короткие ответы на запросы, при этом пользователю не надо звонить и тратить свое время, имеется интуитивно-понятный интерфейс, есть связь с администратором, обслуживает сразу огромную аудиторию пользователей, выдерживая при этом дополнительную нагрузку, обрабатывает большое количество данных. Он надёжен и необходим в век цифровых вызовов и разработок инновационных технологий.

Создание таких сервисов имеет большие перспективы, так как информационная защита клиента в сети – это первостепенная задача, над решением которой работают профессионалы во всем мире. Недаром количе-

ство ботов растет с математической прогрессией. Они неумолимо работают круглые сутки, постоянно улучшается их функционал, ими пользуются в быту и на предприятиях.

Разработанный сервис предоставляет новые возможности для бизнеса. Организациям необходимо сохранять в тайне личные данные своих клиентов, а данный бот может легко проверить эти данные на надежность и предупредить, если есть опасность компрометации учетных записей клиентов компаний, что позволит принять своевременные меры по защите этих данных. Так что за такими ботами будущее.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Горитов А.Н.* Структуры и алгоритмы обработки данных в ЭВМ: Учебное пособие. – Томск: Томский межвузовский центр дистанционного образования, 2000. – 122 с.
2. *Вирт Н.* Алгоритмы и структуры данных. – М.: Мир, 1989. -360 с.
3. *Макконелл Дж.* Основы современных алгоритмов. 2-е дополненное издание.- Москва: Техносфера, 2004. – 368 с.
4. *Ускова О.Ф.* Программирование алгоритмов обработки данных: Учебное пособие. – СПб: БХВ – Петербург, 2003. – 188 с.
5. *Бежанова М.М.* Практическое программирование. Структуры данных и алгоритмы. – М.: Логос, 2001. – 223 с.
6. *Шумилина М.А., Коробко А.В.* Разработка чат-бота на языке программирования PYTHON в мессенджере «TELEGRAM» // Научные известия. - 2022. – С. 47-54.
7. *Карабак И.И., Зорин К.А.* Парсинг TELEGRAM - каналов как элемент системы автоматизированного анализа информации, полученной из сети интернет // Прикаспийский журнал: управление и высокие технологии. – 2022. - № 1(57). – С. 9-17.-DOI: 10.54398/2074-1707_2022_1_9.
8. *Пекшиев Д.В.* Программа для мониторинга и выявления угроз взлома аккаунтов (2023). Свидетельство гос. регистрации программы для ЭВМ № 2023610574.
9. *Пекшиев Д.В.* Программа для обработки и унификации файлов с данными (2023). Свидетельство гос. регистрации программы для ЭВМ № 2023687418.
11. *Пекшиев Д.В.* Программа для индексации набора файлов для бинарного поиска (2023). Свидетельство гос. регистрации программы для ЭВМ. № 2023687348.

Пекишев Даниил Владимирович, аспирант Кубанского государственного университета, Россия, город Краснодар, улица Ставропольская 149, 350040, телефон: +7-900-229-22-65, email: AnaAble1974@gmail.com.

Pekishev Daniil Vladimirovich, graduate student of Kuban State University, Russia, Krasnodar, Stavropolskaya street 149, 350040, phone: +7-900-229-22-65, email: AnaAble1974@gmail.com.