

УДК 004

ТЕЛЕГРАМ-БОТ ДЛЯ ВЫЯВЛЕНИЯ ЭКСТРЕМИСТСКИХ РЕСУРСОВ

Афанасьева А.А., студентка гр. ПИб-211, IV курс

Научный руководитель: Пимонов А.Г., д.т.н., профессор

Прокопенко Е.В., к.ф.-м.н., доцент

Кузбасский государственный технический университет имени Т.Ф. Горбачёва
г. Кемерово

В современном мире экстремизм – это одна из самых острых и актуальных проблем современного общества, затрагивающая различные аспекты жизни и представляющая угрозу как для отдельных людей, так и для целых сообществ. Понимание причин и последствий экстремистских проявлений является важным шагом к их предотвращению и минимизации негативного влияния на общество.

С развитием информационно-коммуникативных технологий и повсеместным распространением Интернета экстремизм приобрел глобальный характер. Виртуальное пространство стало не только платформой для общения и обмена информацией, но и ареной для распространения радикальных идей, вербовки новых последователей и организации насилиственных действий. Проблематика экстремизма в сети Интернет представляет собой сложную задачу, требующую внимания со стороны общества, государства и международных организаций. Для ее решения необходимо разработать комплексные меры, которые могут быть внедрены на разных уровнях: от индивидуального до международного. К таким мерам можно отнести следующие: 1) повышение осведомленности; 2) внедрение систем контроля контента; 3) выявление экстремистских ресурсов и их мониторинг в сети Интернет.

Для повышения осведомленности в ряде стран уже внедрены образовательные программы, направленные на развитие критического мышления у молодежи. Эти программы включают курсы по медиа-грамотности, где обучают распознавать дезинформацию и манипуляции. А для взрослых есть информационные кампании со встречами различных форматов проведения, на которых обучают как поддерживать открытый диалог с молодежью о безопасном поведении в сети.

Важным средством контроля является система контентной фильтрации, блокирующая запрещенный интернет-контент. В соответствии с законодательством Российской Федерации, такая система должна быть установлена в образовательных учреждениях, а также там, где есть публичные точки доступа в интернет, для безопасного использования совместно с механизмом блокировки доступа к фишинговым (мошенническим) сайтам и блокировкой всплывающих окон и баннеров.

Вопрос выявления и мониторинга экстремистских ресурсов в сети Интернет является актуальной задачей для правоохранительных органов, исследователей и общественных организаций. Существует несколько решений, которые помогают в этой области автоматизировать сбор и систематизировать данные: «Сеус», «Open Measures», «Окулус». Но у всех этих инструментов есть одна общая проблема – сложность использования. Их нужно настраивать и разбираться в интерфейсе, что делает их малодоступными для обычных пользователей и усложняет их массовое применение.

Для того чтобы упростить задачу выявления запрещенного контента был разработан Телеграм-бот, задачей которого является осуществление парсинга в сети Интернет. Работает он без сложной установки, имеет интуитивно понятный интерфейс – всё работает прямо в мессенджере, в пару кликов. Это инструмент, которым может воспользоваться любой человек, не имеющий глубоких технических познаний.

Выбор мессенджера Телеграм обусловлен тем, что данная платформа имеет большую популярность среди пользователей, является кроссплатформенной, а также предоставляет удобные возможности для создания чат-ботов. В качестве программной оболочки выбран язык программирования Python, который обладает широкими возможностями и предлагает мощные библиотеки для работы с API Telegram. При разработке использовались библиотеки:

1) telebot (pyTelegramBotAPI) – упрощает процесс кодирования взаимодействия с ботом, так как библиотека легка в изучении и использовании, имеет широкий набор встроенных функций и обладает обширной документацией [**Ошибка! Источник ссылки не найден.**].

2) Python Requests – позволяет отправлять HTTP-запросы разного уровня сложности веб-сервисам и получать от них соответствующие ответы [**Ошибка! Источник ссылки не найден.**].

3) Googlesearch – автоматизирует поиск в Google [1].

4) BeautifulSoup – библиотека для разбора HTML и XML документов, используется для веб-скрейпинга [**Ошибка! Источник ссылки не найден.**].

При работе с Телеграм-ботом доступны следующие команды (**Ошибка! Источник ссылки не найден.**): поиск контента как на конкретном сайте, так и в сети Интернет; кабинет оператора для удобного взаимодействия с результатами, которые сохраняются в таблицу базы данных после поиска; виртуальный список – это один из методов поиска, который позволяет пользователю легко его редактировать, при этом все содержимое также сохраняется уже в другую таблицу.

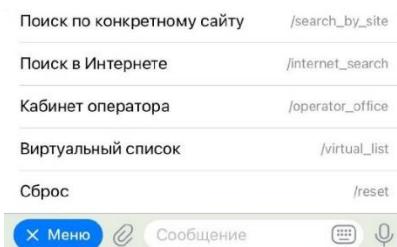


Рисунок 1 – Команды Телеграм-бота

Для демонстрации работы поиск будет осуществлен в интернете по списку, содержащему формы экстремизма. После выбора команды «Поиск в Интернете», нужно выбрать одну из команд способов поиска (рис. 2). В данном случае выбираем «Загрузить файл со списком» и отправляем файл Телеграм-боту. При необходимости пользователь может, не выходя из чата, отредактировать список и после этого начать по нему поиск.

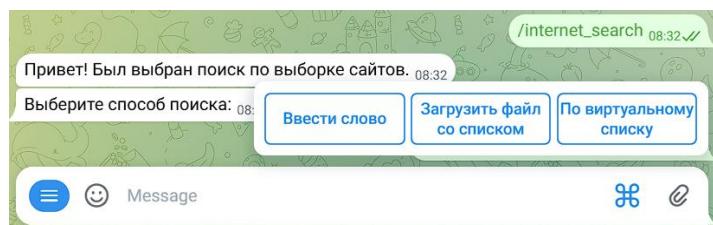


Рисунок 2 – Команды способов поиска

Для понимания того, на какой стадии поиска находится Телеграм-бот, в чат отправляется сообщение по мере завершения анализа страниц по какому-либо слову. После окончания работы пользователю отправляются краткие итоги поиска (рис. 3): количество ресурсов, которые были проанализированы, общее число ссылок, на страницах которых обнаружены искомые слова и, при наличии, число ошибок, возникших при обработке ресурсов. А также отправляется статистика, показывающая число веб-страниц, содержащих конкретное слово.

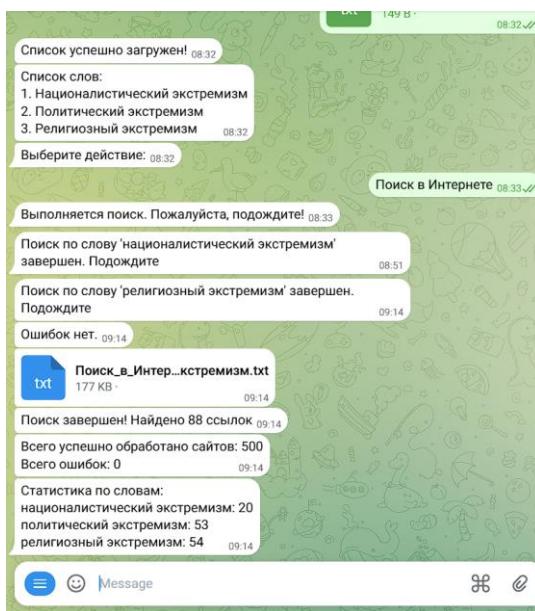


Рисунок 3 – Результаты работы

Помимо этого пользователю отправляется файл, содержащий отфильтрованные по словам предложения и веб-адреса, на которых они найдены (рис. 4). Также если при обработке сайта возникают ошибки, то причина ошибки и

адрес сайта записываются в отдельный файл и таким же образом отправляются пользователю.

Поиск в Интернете списка из 3 слов: националистический, экстремизм – блокнот
Файл Правка Формат Вид Справка
Результат поиска по словам националистический экстремизм, политический экстремизм, религиозный экстремизм:

1. ЧТО ТАКОЕ ЭКСТРЕМИЗМ? | Официальный сайт Администрации Металлургического района города Челябинска
https://metchel.gov74.ru/Publications/terror_news>Show?id=6081

НАЦИОНАЛИСТИЧЕСКИЙ ЭКСТРЕМИЗМ
1. националистический экстремизм. национализм отвергает интересы и права других наций, провозглашав верховенство

ПОЛИТИЧЕСКИЙ ЭКСТРЕМИЗМ
1. политический экстремизм – является самым распространенным
2. религиозно-политический экстремизм – это проявление политической деятельности, закамуфлированной под религиозные

Количество найденных предложений: 3

2. Экстремизм – Википедия
<https://ru.wikipedia.org/wiki/%D0%AD%D0%BA%D1%81%D1%82%D1%80%D0%B5%D0%BC%D0%BB%D0%B7%D0%BC>

ПОЛИТИЧЕСКИЙ ЭКСТРЕМИЗМ
1. религиозно-политический экстремизм
РЕЛИГИОЗНЫЙ ЭКСТРЕМИЗМ
1. религиозный экстремизм
2. религиозный экстремизм в современной россии: социально-философский аспект // проблемы противодействия экстремизму
3. религиозный экстремизм: сущность, причины, пути преодоления / автореферат диссертации

Количество найденных предложений: 4

3. Профилактика экстремизма и терроризма - Муниципальные новости - Новости, объявления, события - Администрация
<https://adm-sosva.ru/news/media/2019/12/5/profilaktika-ekstremizma-i-terrorizma/>

НАЦИОНАЛИСТИЧЕСКИЙ ЭКСТРЕМИЗМ
1. националистический экстремизм – это насилие по мотивам национальной или расовой ненависти и вражды, возбуждение

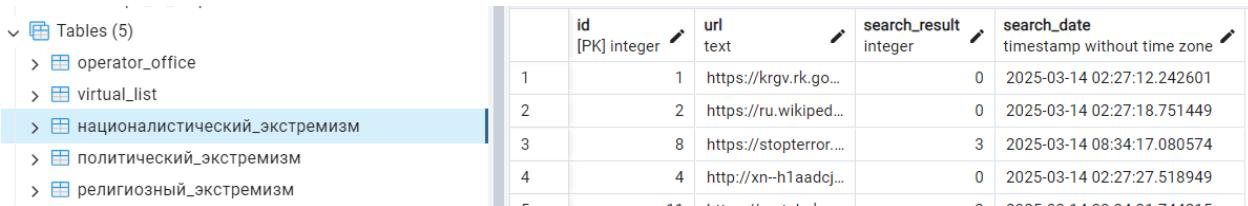
ПОЛИТИЧЕСКИЙ ЭКСТРЕМИЗМ
1. политический экстремизм – теория и практика насилиственного, незаконного взятия и удержания государственно

РЕЛИГИОЗНЫЙ ЭКСТРЕМИЗМ
1. религиозный экстремизм – практика насилиственного насилия религиозной веры среди неверующих либо представителей

Количество найденных предложений: 3

Рисунок 4 – Содержимое файла с результатами поиска

Для того чтобы избежать такой ситуации, что при повторном поиске будут найдены и отправлены пользователю те же самые ссылки, что и при первом поиске, по каждому слову в СУБД PostgreSQL создается таблица, в которой хранится ссылка на ресурс, количество найденных предложений по этому слову и время добавления (рис. 5).



	id [PK] integer	url text	search_result integer	search_date timestamp without time zone
1	1	Show?id=6081">https://krvg.rk.gov74.ru/Publications/terror_news>Show?id=6081	0	2025-03-14 02:27:12.242601
2	2	https://ru.wikipedia.org/wiki/%D0%AD%D0%BA%D1%81%D1%82%D1%80%D0%B5%D0%BC%D0%BB%D0%B7%D0%BC	0	2025-03-14 02:27:18.751449
3	8	https://stopterror.ru/	3	2025-03-14 08:34:17.080574
4	4	http://xn--h1aadcj1a.xn--p1ai/	0	2025-03-14 02:27:27.518949

Рисунок 5 – Таблица «Националистический_экстремизм»

Можно сделать вывод, что использование Телеграм-бота позволяет существенно экономить самый ценный человеческий ресурс – время. По результатам демонстрации видно, что за 41 минуту было проанализировано 500 сайтов по нужным пользователю словам, что впечатляет, поскольку человеку было бы сложно просмотреть такое количество ресурсов за такой короткий промежуток времени. Помимо этого, использование разработанного инструмента на такой популярной и удобной платформе станет доступным для людей без специальных знаний, что может привести к его широкому распространению. Простота, удобство и доступность – вот что действительно имеет значение в современных условиях. И важно отметить, что с помощью созданного бота можно эффективно искать запрещенный контент и отправлять его на блокировку. Это не только поможет поддержать порядок в интернете, но и способствует созданию более безопасной онлайн-среды для всех пользователей.

Таким образом, использование Телеграм-бота не только является доступным и удобным, но и играет значимую роль в повышении безопасности в сети.

Список литературы:

1. Обзор возможностей библиотеки telebot для создания Telegram ботов.
– URL: <https://skyeng.ru/it-industry/programming/biblioteka-telebot-dlya-sozdaniya-telegram-botov/?ysclid=m87gt8awm3753410998> (дата обращения: 09.03.2025).
2. Python Requests: что это за библиотека и как ее использовать. – URL: <https://blog.skillfactory.ru/python-requests/> (дата обращения: 09.03.2025).
3. Библиотека для создания идеального скрипта автоматизации поиска на Python. – URL: <https://dzen.ru/a/YpRN6ZDnCGHdtZPR?ysclid=m87ihc0cm5946234035> (дата обращения: 09.03.2025).
4. Веб-скраппинг с помощью BeautifulSoup – пошаговое руководство. – URL: <https://python-scripts.com/beautifulsoup-parsing> (дата обращения: 09.03.2025).