

АВТОМАТИЗАЦИЯ АЛГОРИТМА РАБОТЫ С ОТКРЫТЫМИ ИСТОЧНИКАМИ ИНФОРМАЦИИ В СОВРЕМЕННЫХ УСЛОВИЯХ

*Макаров И.А., студент группы СГНЗ-63Б
Московский государственный технический университет им. Н.Э. Баумана
ivanmakarovxy@gmail.com*

*Научный руководитель: Муравьева Е.О., старший преподаватель
кафедры СГН-3 «Информационная аналитика и политические технологии»
muraviova@bmstu.ru*

Аннотация: в работе рассматривается актуальная проблематика современной повестки дня: открытые источники и специфика их использования в современных условиях. В исследовании открытые источники информации определяются как элементы в системе информационных ресурсов, предоставляющие открытую информацию и возможность ее свободного использования, модификации и распространения.

Ключевые слова: открытые источники информации, виды открытых источников информации, социальные сети, алгоритм работы с открытыми источниками информации, автоматизация алгоритма работы с открытыми источниками информации.

В современном мире информация становится основополагающей силой, определяющей направления развития технологий, экономической жизни и культурных процессов. Источники информации являются критически важными элементами этого развития, формируя фундамент, на котором строится эта сила. Источники информации делятся на: открытые, условно-открытые и закрытые [10]. Каждый вид имеет свою специфику, но в рамках этой статьи, мы будем фокусироваться на открытых источниках. Открытые источники информации можно определить как элементы в системе информационных ресурсов, предоставляющие открытую информацию и возможность ее свободного использования, модификации и распространения [10].

В этой статье мы рассмотрим основные виды открытых источников информации, определим наиболее популярные и востребованные типы в современных условиях и их основные характеристики, а также выделим алгоритм работы с открытыми источниками, проблемы при реализации этого алгоритма и выявим возможности автоматизации его этапов.

Рассмотрев широкий спектр источников, начиная от всемирной сети интернет с её порталами, социальными сетями и блогами, через СМИ в их многочисленных форматах, до официальных отчетов компаний и профессиональных сетевых мероприятий [10]. Можно прийти к выводу о том, что каждый источник имеет свою уникальную роль в распространении информации и может использоваться для различных

целей и задач. Основываясь на существующих видах открытых источников, можно сформировать ментальную карту, описывающую структуру открытых источников информации, представленную на Рис. 1.

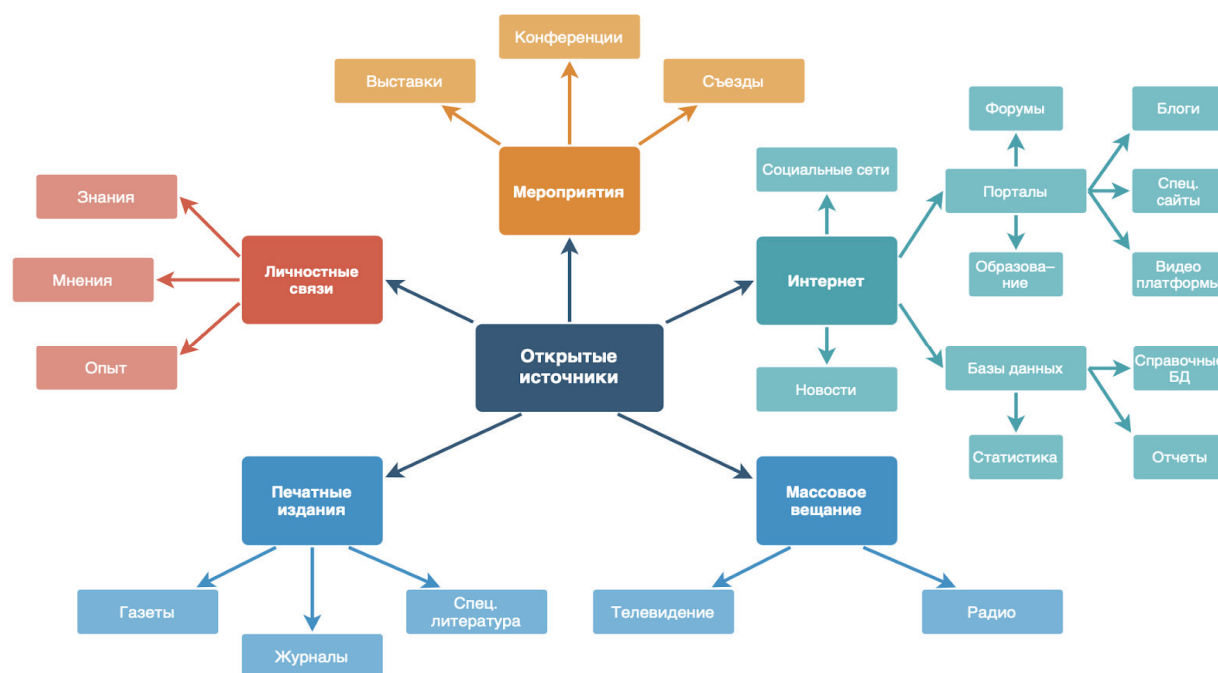


Рис. 1. Структура открытых источников

Основные свойства открытых источников [4]: доступность, гибкость, прозрачность, многообразие. Ценность сведений, получаемых из открытых источников: актуальность, репрезентативность, свободное распространение, отсутствие репутационных рисков.

На сегодняшний день наиболее актуальным является изучение открытых источников в интернете. Все аналитические задания становятся чрезвычайно сложными без доступа к Интернету, в современном мире проведение отдельных исследований без использования информационных технологий и ресурсов сети просто невозможно [10]. Интернет предлагает для исследования практически любых процессов такие возможности, какие не предоставляет никакой другой источник информации.

Если обратиться к глобальной статистике, то по последним данным ITU и GSMA Intelligence [11], количество пользователей интернета в мире по состоянию на 2023 год составляет 5,16 миллиарда. Это значит, что 64,4% мирового населения имеют доступ в интернет. За год количество интернет-пользователей выросло на 1,9%. Если говорить про популярность использования интернета в России, то здесь интернет также занимает лидирующие позиции. Так, доля пользователей интернета в возрасте от 12 до 24 лет в России приблизилась к 100% и составила 97,1% за февраль – ноябрь 2023 года, следует из данных Mediascope [9].

По данным GSMA Intelligence [2] за 2023 год из топ-7 самых популярных интернет-ресурсов во всем мире, больше половины приходится на социальные сети. Если рассматривать популярность интернет-ресурсов в России за последнее время, то можно прийти к похожим выводам. Так, по данным Mediascope [9] за 2023 в топ-7 самых популярных интернет-порталов входят 4 социальные сети. Самой широко используемой на данный момент социальной сетью в России является Telegram [9]. Telegram – это одно из наиболее часто используемых приложений для обмена мгновенными сообщениями во всем мире [12].

Таким образом, исходя из всего вышеперечисленного, мы можем сделать вывод о том, что интернет на сегодняшний день является самым актуальным, популярным и используемым ресурсом для получения информации как в России, так и во всем мире. Самым популярным видом открытых источников в интернете являются социальные сети – это явление также характерно как для России, так и для всего мира в целом. А самой популярной социальной сетью является Telegram. В связи с этим в дальнейшей работе мы сфокусируемся на анализе социальной сети Telegram и под социальными сетями будем иметь в виду эту социальную сеть.

К уникальным особенностям социальных сетей можно отнести [1]: простота публикации контента, анонимность и прозрачность, широкий охват пользователей, независимый информационный поток, симбиоз между традиционными СМИ, независимым информационным потоком рядовых пользователей, а также пространством сети Интернет, что и создает уникальную ценность для работы аналитика с этим видом источников.

Выделяют следующий алгоритм работы с открытыми источниками информации [4]:

1. Целеуказание и планирование
2. Сбор данных
3. Обработка данных
4. Анализ и синтез информации
5. Распространение
6. Алгоритм является циклическим и каждый его этап включает

в себя дополнительные составляющие.

Можно выделить следующие проблемы открытых источников и трудности реализации приведенного алгоритма работы с открытыми источниками:

Большой объем и разнообразие как открытых источников информации, так и самой информации, которую они содержат [3]. Разнообразие и объем открытых источников информации затрудняют этап сбора данных, требуя от аналитика определения наиболее релевантных источ-

ников и информации для анализа. Это также повышает сложность обработки данных, так как необходимо эффективно фильтровать и структурировать большой объем информации.

Низкая надежность открытых источников [4]. На этапе сбора данных аналитику приходится оценивать и выбирать источники с высокой степенью надежности, чтобы избежать использования ненадежной информации. На этапе обработки данных требуется проводить дополнительную фильтрацию и проверку информации на достоверность. Эти факторы также влияют на анализ и синтез данных, где необходимо учитывать возможные искажения и ошибки при формировании выводов и рекомендаций.

Низкая достоверность информации, находящейся в открытых источниках [4]. На алгоритм анализа открытых источников низкая достоверность информации влияет, прежде всего, на этапы сбора и обработки данных.

Необходимость фильтрации и проверки открытых источников [5]. Является основной проблемой при работе с открытыми источниками информации. Сильно увеличивает время информационно-аналитической работы, особенно на этапах планирования и сбора данных.

В связи с этим, нашей дальнейшей задачей будет решение основной проблемы работы с открытыми источниками с помощью современных технологий работы с открытыми источниками: необходимость фильтрации и проверки открытых источников. Разработав инструмент для проверки и фильтрации открытых источников информации, а именно социальных сообществ, где и публикуется основной контент, можно автоматизировать алгоритм работы с открытыми источниками, а именно этап № 2 – сбор данных, повысив тем самым скорость проведения информационно-аналитической работы и эффективность достижения поставленных целей работы аналитика.

Разработанный инструмент предоставляет из себя телеграмм-бот для проверки и фильтрации сообществ в социальной сети Telegram (телеграмм каналов). Телеграмм-бот имеет следующие назначение: принимая на вход от пользователя ссылку на телеграмм канал (далее тг-канал), бот предоставляет меню действий с этим каналом для просмотра статистики и аналитики канала. Телеграмм-бот написан на языке программирования Python, получает данные с помощью API-запросов к Telegram и сервису для аналитики телеграмм-каналов TGstat, использует базу данных MySQL для хранения полученных данных.

Рассмотрим функциональные требования разрабатываемого инструмента. После отправки тг-канала, у пользователя появляется меню с кнопками для вызова действий:

Сводная статистика тг-канала: (название, географическое положение, язык, категория, дата создания, возраст канала, аватарка канала)

Статистика аудитории и просмотров тг-канала (число подписчиков, динамика изменения подписчиков (в виде графика), динамика изменения просмотров (в виде графика), вовлечение (ER) – какое количество пользователей активно взаимодействует с контентом сообществ, вовлечение на основе охвата (ERR).

Все измерения, отображаемые в виде графика, по умолчанию отображаются за все время, пользователь может изменить отображение динамики по дням/месяцам/годам с помощью кнопок под графиком.

Статистика публикаций тг-канала (динамика изменения числа публикаций (в виде графика), динамика изменения числа репостов публикаций (в виде графика), цитируемые и цитирующие тг-каналы, число цитируемых тг-каналов и число цитирующих тг-каналов выбранный тг-канал.

Анализ тг-канала (проверка не является ли канал отмеченным как мошеннический/фейковый/накрученный и т.д.). Проверка происходит с помощью поиска упоминания тг-канала в базе данных Telegram и TGstat.

По результатам проверок выводится статус проверки с указанием проверяемых величин. Статуса может быть 2: «Зеленый» – все проверки пройдены успешно; «Красный» – канал упоминается как мошеннический/фейковый/накрученный и т.д.

Таким образом, разработанный инструмент позволит аналитику с помощью современных технологий производить проверку и фильтрацию телеграмм каналов для дальнейшего использования в рамках алгоритма работы с открытыми источниками информации. С помощью данного инструмента автоматизируется алгоритм работы с открытыми источниками информации, тем самым повышается скорость проведения информационно-аналитической работы и эффективность достижения поставленных целей работы аналитика.

Литература и источники

1. Браславец Л.А. Интернет-сервисы социальных сетей в современной системе средств массовой информации. Дисс. ... канд. фил. наук. Воронеж, 2020.
2. Бочарников И.В. Террористическая группировка «Исламское государство» как закономерное следствие ближневосточной стратегии США // Тренды и управление. 2015. № 4. С. 395–402.
3. Демидов В.В. Информационно-аналитическая работа в международных отношениях: Учебное пособие. – Новосибирск: НГАЭиУ, 2019. 192 с.
4. Доронин А.И. Бизнес-разведка 2.2 + OSINT. – 7-е изд., перераб. и доп. – М.: ДМК Пресс, 2023.
5. Курносков Ю.В., Конотопов П.Ю. Аналитика: методология, технология и организация информационно-аналитической работы. М.: «Русаки», 2004.
6. Муравьева Е.О. Геополитические условия реализации интересов России на пространстве бывшей Югославии. В сборнике: Мегатренды мировой поли-

тики: глобализация, поляризация, экстремизм. Сборник материалов Международной научно-практической конференции. Москва, 2023. С. 377–386.

7. Постсоветское пространство 30 лет спустя / Кошкин А.П., Бочарников И.В., Манойло А.В., Бондалетов В.В., Глазунов О.Н., Давыдова Ю.А., Денисенкова Н.Н., Масликов В.А., Новиков А.В., Овсянникова О.А., Перенджиев А.Н., Севостьянов П.И., Черданцев В.В. Москва, 2021.

8. Ремарчук В.Н. Глобализация и ее влияние на российско-американские отношения // Политика и общество. 2015. № 3 (123). С. 407–413.

9. Технологичная исследовательская компания Mediascope. <https://mediascope.net/>.

10. Унижаев Н.В. Информационно-аналитическое обеспечение безопасности организации: для вузов и для дополнительного образования. СПб: Интермедия, 2018. 408 с.

11. GSMA Intelligence, Digital 2023: Global overview report. <https://datareportal.com/reports/digital-2023-global-overview-report>.

12. La Morgia M. et al. It's a Trap! Detection and Analysis of Fake Channels on Telegram //2023 IEEE International Conference on Web Services.

13. La Morgia M. et al. uncovering the dark side of telegram: Fakes, clones, scams, and conspiracy movements. 2021. С. 1–12.