С. А. Молоков, С. А. Рожков, Т. И. Паюсова
Тюменский государственный университет, г. Тюмень
УДК 004.056

ФИЛЬТРАЦИЯ МЕДИАКОНТЕНТА ВЕБ-СТРАНИЦ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация: В статье рассмотрена проблема фильтрации нежелательного контента. Описан целостный подход для автоматизированного обнаружения, распознавания и фильтрации запрещенного медиаконтента с использованием нейросетевого моделирования.

Ключевые слова: информационная безопасность, медиаконтент, фильтрация медиаконтента, нейронные сети.

Введение. В наши дни использование Интернета стало неотъемлемой частью повседневной жизни. Параллельно активному росту публикаций информации в сети, растет и распространение запрещенного контента в Интернете, вместе с этим появляется необходимость в разработке методов пресечения распространения такого контента [1, 2].

Мы все чаще можем наткнуться на запрещенный контент на вебстраницах, вопреки действиям, предпринимаемым владельцами вебресурсов и правительством РФ, для борьбы с этим явлением. Так, например, согласно п. 5, ч.1, ст. 10.6, ФЗ 149 «Об информации, информационных технологиях и о защите информации» [3] Владелец страницы сайта в сети "Интернет" обязан осуществлять мониторинг информационного ресурса в целях выявления запрещенного контента, в частности, содержащего: способы приготовления и использования наркотических веществ (ответственность ст. 228.1 УК РФ [4]), способы совершения и призывы к суициду (ответственность ст. 110, ст. 110.1, ст. 110.2 УК РФ), оскорбление и агрессию в сторону других граждан, государства и государственной символики (ответственность ст. 282, ст. 329 УК РФ), информацию пропагандирующую ЛГБТ (ответственность ст. 6.21 КоАП РФ[5]), призывы к экстремистской и террористической деятельности(ответственность ст. 205.2, ст. 280 УК РФ) и т. д.

Проблема исследования. Согласно статистике ВКонтакте за 2022 г. было опубликовано 6.3 млрд ед. контента [6]. Путем несложных

математических вычислений можно подсчитать, что в среднем публикуется ~ 17 260 274 ед. контента/ в день, это ~ 719 178 ед. контента/ в час, это ~ 11 986 ед. контента/ в минуту, или же около 200 записей в секунду (и это лишь в самой крупной социальной сети в РФ).

При таком объеме поступления информации на онлайн-платформы не представляется возможным ручная фильтрация всех материалов на соответствие этическим нормам и законодательным требованиям. Можно сделать вывод, что со временем, при активном росте количества информации, будет увеличиваться распространение вредоносного и нежелательного контента, который будет пагубно влиять на пользователей и сам веб-ресурс.

Таким образом, целью нашей работы является разработка собственного комплексного подхода для автоматизированного обнаружения, распознавания и фильтрации медиаконтента с использованием нейронных сетей.

Материалы и методы.

Сравнительный анализ библиотек для парсинга веб-страниц.

Парсинг веб-страниц — автоматическое извлечение и систематизация данных с веб-страницы. Чаще всего процесс состоит из анализа HTML-кода страницы, для получения такой информации как текст, изображения, видео, аудио и т. д. [7]. В нашей работе мы рассмотрим и сравним несколько общедоступных библиотек для парсинга: BeautifulSoup, Selenium, Ixml, Scrapy [8]. Результаты сравнения библиотек представлены в виде таблицы (табл. 1).

Таблица 1

Сравнительная таблица библиотек для парсинга веб-страниц

Библиотека	*Язык программирования*	*Парсинг JavaScript*	*Обработка динамических страниц*	*Скорость работы*	*Лицензия*
Beautiful Soup	Python	-	-	3	MIT
Ixml	Python	-	-	2	BSD
Scrapy	Python	-	+	1	BSD
Selenium	Java, Ruby, C#, Python, Kotlin и др.	+	+	4	Apache

Таким образом, мы сделали вывод о том, что Scrapy лучше подходит для парсинга статичных веб-страниц, в то время как Selenium способен обрабатывать и динамические страницы в том числе, а также эффективнее справляется с парсингом JavaScript, однако скорость его работы заметно ниже.

Классификация медиаконтента.

Для того, чтобы эффективно фильтровать медиаконтент, необходимо иметь в виду то, что контент может быть разнообразным, и, в зависимости от вида контента, требуется использовать различные методы фильтрации. В связи с этим мы предлагаем проводить классификацию медиаконтента на текст и изображения, так как для фильтрации этих групп необходимо использовать разные подходы.

К группе текстовых данных автоматически примыкает текст на веб-страницах, также в эту группу можно отнести аудиальный контент, такой как голосовые сообщения, музыка и т. п., который может быть расшифрован и преобразован в текст, для этого можно применить библиотеку Vosk для Python [9], и в дальнейшем производить фильтрацию как текстовых данных.

Группу изображений можно представить как изображения, видео, а также Gif-анимации. Видеоизображения могут быть рассмотрены как поток кадров и аудиодорожка, которую можно извлечь с помощью библиотеки FFmpeg для Python и представить в виде текста, используя уже упомянутую библиотеку Vosk также для Python. Для того, чтобы разбить видео и Gif-анимации на кадры можно воспользоваться библиотекой OpenCV на Python [10]. Таким образом, видеоконтент может быть отнесен сразу к двум типам контента: текстовые данные и изображения. Полная модель классификации представлена на рис. 1.

Можно сделать вывод, что классификация медиаконтента упростит его фильтрацию в дальнейшем, и, в зависимости от вида контента, позволит на следующем этапе использовать всего два подхода: фильтрация текста и фильтрация изображений.

Применение нейронных сетей при принятии решения об отнесении контента к запрещенному.

В наше время нейронные сети уже активно используются в сфере безопасности, они способны обнаруживать и отслеживать объекты на изображениях и видеозаписях, также извлекать целевой атрибут [11].

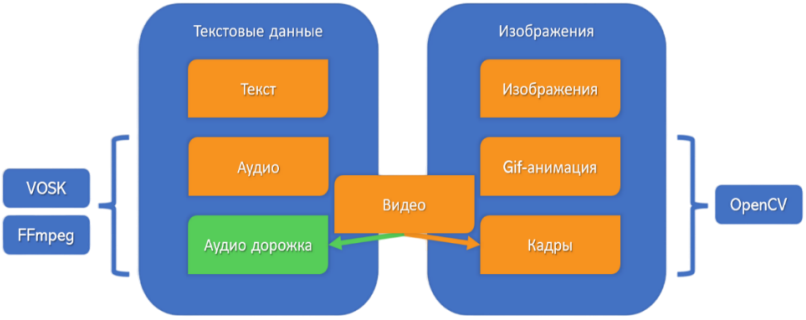

Рис. 1. Модель классификациии медиаконтента

В рамках нашей статьи мы, в первую очередь, решили проанализировать уже существующие готовые решения для фильтрации медиаконтента отечественного производства, так как они создавались в соответствии с требованиями Российского законодательства.

Нам удалось найти лишь два решения: это нейросеть от ВКонтакте [12] и “Окулус” от Роскомнадзора [13, 14]. При попытке углубиться в детали и методы работы упомянутых нейронных сетей мы столкнулись с проблемой, что данные решения закрыты для глаз простых обывателей, так что нам не удалось напрямую познакомиться с ними. Однако удалось выяснить, что нейросеть “Окулус” создана в целях правового регулирования: для поиска веб ресурсов и конкретных лиц, распространяющих запрещенный контент, а также для блокировки таких ресурсов с дальнейшей передачей материалов в правоохранительные органы.

По причине недоступности упомянутых выше отечественных нейросетей мы отправились на поиск решений от наших зарубежных коллег. В результате поиска нам удалось выбрать несколько нейросетей подходящих для наших задач, выделим две из них:

1) Нейросеть для фильтрации изображений — Google Cloud Vision API — это нейросеть, способная распознавать текст, лица, логотипы, классифицировать изображения и выявлять часть запрещенного контента на них [15]. Возможности Google Cloud Vision API: считывание печатного и рукописного текста на изображениях; распознавание образов на изображениях; распознавание нежелательного контента на изображениях; классификация изображений по миллионам предопределенных категорий.

2) Нейросеть для фильтрации текстовых данных — Google Cloud Natural Language API, которая способна выполнять следующие действия [16]: синтаксический анализ; анализ настроений; анализ сущностей; анализ настроений сущностей; текстовая классификация.

Однако стоит учесть, что, во-первых, использование предложенных решений, хоть они и имеют бесплатный пробный период, предусматривает плату в зависимости от масштабов использования; во-вторых, нет полной уверенности в том, что данные решения смогут полностью покрыть необходимый нам спектр задач по фильтрации запрещенного контента, в силу того, что они не были специально обучены под фильтрацию контента, обозначенного законодательством РФ как запрещенный.

В связи с данными ограничениями мы пришли к выводу, что самым оптимальным вариантом является разработка и обучение собственной нейросети для дальнейшего ее использования в приложении для автоматической фильтрации медиаконтента веб-ресурсов на основе предложенного нами алгоритма.

Результаты. Для достижения поставленной цели были проанализированы существующие решения в данной предметной области, разработан комплексный подход для автоматизированного обнаружения, распознавания и фильтрации медиаконтента с использованием нейронных сетей (рис. 2).

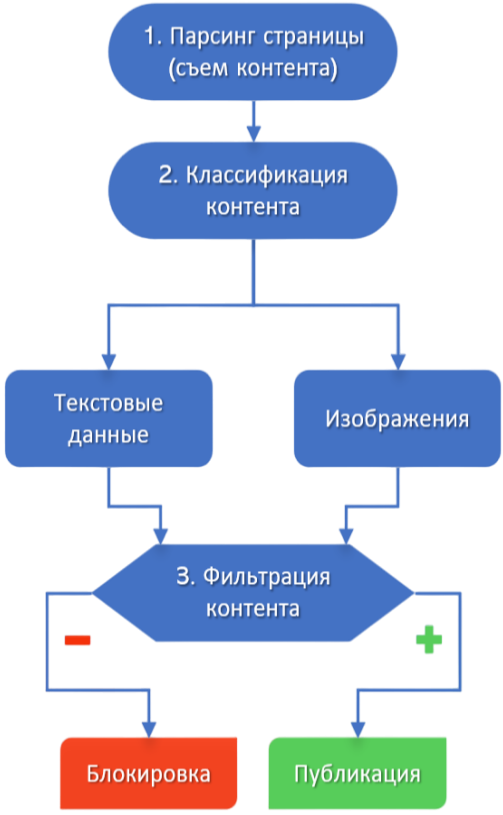

Рис. 2. Схема предложенного алгоритма

Заключение. Используя текущие теоретические наработки в будущих исследованиях, мы планируем реализовать на практике обучение собственной нейронной сети для распознавания и фильтрации медиаконтента и, в дальнейшем, разработать собственный продукт для автоматической фильтрации контента на веб-страницах, с открытым исходным кодом и предоставить его бесплатно всем желающим.

СПИСОК ЛИТЕРАТУРЫ

1. Фролов А. А. Анализ механизмов обнаружения запрещенного содержимого в сети Интернет / А. А. Фролов, Д. С. Сильнов, А. М. Садретдинов. — Текст : непосредственный // International Journal of Open Information Technologies. — 2019. — № 1. — С. 90-96.
2. Балашов А. Н. Правовое регулирование Интернет-отношений: основные проблемы и практика реализации в России / А. Н. Балашов. — Текст : непосредственный // Среднерусский вестник общественных наук. — 2016. — № 2. — С. 113-118.
3. Об информации, информационных технологиях и о защите информации: Федеральный закон № 149-ФЗ: принят Государственной Думой 08.07.2006 г.: одобрен Советом Федерации 14.07.2006 г. — Москва, Кремль, 2006. — Текст : непосредственный.
4. Российская Федерация. Законы. Уголовный кодекс Российской Федерации : УК : текст с изм. и доп. на 28.04.2023 г. — Москва : Эксмо, 2023. — Текст : непосредственный.
5. Российская Федерация. Законы. Кодекс Российской Федерации об административных правонарушениях:КоАП: : текст с изм. и доп. на 17.05.2023 г. — Москва, 2023. — Текст : непосредственный.
6. Годовой отчет VK за 2022 год. — Текст : электронный // VK:[сайт].– URL: https://corp.vkcdn.ru/media/files/vkarfy2022ru_Va71SXH.pdf (дата обращения 25.04.2023)
7. Просветов В. Л. Анализ методов и средств автоматизации процессов обработки данных веб-сайтов / В. Л. Просветов, Н. Е. Конева. — Текст : непосредственный // Евразийское Научное Объединение. — 2019. — № 1-2. — С. 89-94.
8. Москаленко А. А. Разработка приложения веб-скрапинга с возможностями обхода блокировок / А. А. Москаленко, О. Р. Лапонина, В. А. Сухомлин. — Текст : непосредственный // Современные информационные технологии и ИТ-образование. — 2019. — № 2. — С. 413-420.
9. Самигулин Т. Р. Определение маркеров агрессивного поведения человека на основе анализа аудио и текстового каналов / Т. Р. Самигулин, И. З. Смирнов, А. А. Лаушкина. — Текст : непосредственный // Научный результат. Информационные технологии. — 2022. — № 2. — С. 55-62.
10. Магамедова Д. М. OpenCV-инструмент компьютерного зрения / Д. М. Магамедова. — Текст : непосредственный // Тенденции развития науки и образования. — 2020. — № 63-3. — С. 42-48.

11. Панов А. И. Анализ применения искусственного интеллекта в сфере безопасности / А. И. Панов. — Текст : непосредственный // Экономика и качество систем связи. — 2022. — № 4(26). — С. 46-53.
12. Нейросеть «ВКонтакте» научили выявлять суицидальный контент. — Текст : электронный // Известия : [сайт]. — URL: https://iz.ru/news/673264 (дата обращения 30.04.2023).
13. Роскомнадзор планирует использовать искусственный интеллект для борьбы с дезинформацией в интернете. — Текст : электронный // Комсомольская правда : [сайт]. — URL: https://www.kp.ru/online/news/5240932/ (дата обращения 27.04.2023).
14. «Окулус» запущен: теперь за вашими фото и видео следят. — Текст : электронный // TexTerra : [сайт]. — URL: https://texterra.ru/blog/okulus-roskomnadzora-nachal-iskat-zapreshchennye-materialy-v-internete.html (дата обращения 30.04.2023).
15. Hosseini H. Google's Cloud Vision API is Not Robust to Noise / H. Hosseini, Xiao,and,R B. — Текст : непосредственный // 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). — Mexico : IEEE, 2017. — С. 101-105.
16. NLP With Google Cloud Natural Language API. — Текст : электронный // TopTal : [сайт]. — URL: https://www.toptal.com/machine-learning/google-nlp-tutorial (дата обращения 04.05.2023).