Инженерный вестник Дона, №9 (2024)
ivdon.ru/ru/magazine/archive/n9y2024/9479

Экспертная система регулирования доступа к деструктивным интернет-ресурсам

А.А. Джуров, М.С. Сердечный, Л.В. Черкесова, Е.А. Ревякина

Донской Государственный Технический Университет, г. Ростов-на-Дону

Аннотация: В настоящее время интернет стал неотъемлемой частью нашей жизни, предоставляя доступ к огромному количеству информации и сервисов. Однако, вместе с этим, растет и количество деструктивных интернет-ресурсов, которые могут нанести вред пользователям, особенно детям и подросткам. В связи с этим, возникает необходимость создания эффективной системы регулирования доступа к таким ресурсам. В статье представлена экспертная система регулирования доступа к деструктивным интернет-ресурсам, разработанная на основе современных технологий и методов искусственного интеллекта. Система позволяет автоматически выявлять и блокировать доступ к ресурсам, содержащим вредоносный контент, а также предоставляет возможность для ручной настройки и контроля доступа. В статье описаны основные компоненты системы, а также представлены изображения, демонстрирующие работу системы для блокирования доступа к деструктивным ресурсам. Статья будет полезна для специалистов в области информационной безопасности, искусственного интеллекта и защиты детей от вредоносного контента в интернете.

Ключевые слова: деструктивный контент, экспертная система, информационная безопасность, интернет-ресурсы, SpaCy, Keras, RNN, LSTM, PyQt5, векторизация.

Введение.

В России количество пользователей на момент того же 2024 года достигло 130,4 миллионов пользователей, согласно ежегодному отчету Digital от Datareportal, We Are Social и Meltwater [1].

Такие показатели объясняются широким распространением устройств, имеющих в своем функционале доступ в интернет, таких, как смартфоны, персональные компьютеры, большинство моделей современных телевизоров и т.д.

Одним из рисков интернета и его широкого распространения – это деструктивный контент. К данному типу контента относится информация, которая несет угрозу здоровью (физическому или психическому) и жизни как самого пользователя, так и других людей. Согласно новостному сайту «Известия», за 2023 год Роскомнадзор заблокировал более 31,5 тысяч материалов с деструктивным контентом на основании обращений

Росмолодежи, занимающейся мониторингом интернет-пространства на наличие деструктивных интернет-ресурсов. В 2022 году данных показатель составил 18,7 тысяч блокировок [2].

Вышеописанные тенденции сделали очень актуальными разработки в сфере автоматизации контентного регулирования в интернете. В связи с чем разработки в этом направлении ведутся как в частных компаниях, так и государственных структурах.

Государства в целях регулирования запрещенного контента могут привлекать к ответственности владельцев сайтов, на которых размещен такой контент или требовать от них удалять информацию, противоречащую законодательству конкретного государства. Или же у государственных организаций могут быть инструменты, обеспечивающие блокировку запрещенных ресурсов, например, Роскомнадзор выдает операторам связи клиент автоматической системы, который осуществляет блокировку IP-адресов запрещенных сайтов. Кроме того, у операторов могут быть свои базы данных заблокированных ресурсов.

Проверкой контента занимаются как государственные структуры (Роскомнадзор), так и частные лица, например, в России существует ряд общественных организаций, занимающихся анализом контента в интернете и передачей данных о выявленных нарушениях в соответствующие государственные органы.

Владельцы интернет-ресурсов также занимаются модерацией контента, они могут использовать для этого несколько подходов:

– Модератор-человек, регулированием контента на конкретном сайте занимается отдельный специалист – модератор, который основываясь на своем субъективном заключении может удалить пост, статью, комментарий и т.д. Данный подход является самым простым в реализации, однако

эффективность его уменьшается при увеличении аудитории и публикуемого контента.

– Модераторы-пользователи, при помощи функционала сайта пользователи отправляют жалобы на конкретный материал размещенный на сайте. Из-за большого количества недостатков, почти неэффективен как самостоятельный инструмент: пользователи могут пренебрегать или наоборот злоупотреблять подобной функцией. Кроме того, сам подход предполагает, что некоторые пользователи обязательно должны столкнуться с деструктивным контентом, в то время как основная задача модерации – по возможности полностью исключить подобную ситуацию.

– Программные средства регулирования контента. Данный подход подразумевает использование как относительно простых методов таких как списки запрещенных фраз и выражений, так и более продвинутых, с применением технологий искусственного интеллекта. Использование подобного подхода, позволяет увеличить скорость проверки контента и оптимизировать штат модераторов, однако главным минусом этого подхода является относительно низкая точность работы, компьютер часто неспособен верно интерпретировать контекст из-за чего может пропустить запрещенный пост или статью, или же вовсе заблокировать публикацию, в которой нет запрещенной информации [3].

– Комбинированный подход, является самым широко используемым среди частных интернет-ресурсов. Основная идея – максимально уменьшить влияние недостатков вышеописанных подходов на эффективность работы системы за счет достоинств друг друга.

По мимо всего вышеописанного регулированием доступу к интернет-ресурсам может заняться конкретный пользователь, например, с целью осуществить родительский контроль над ребенком. Достигается это путем создания черного списка IP- или URL-адресов при помощи функционалов

операционных систем, маршрутизаторов или программ, предоставляющих доступ в интернет, таких как браузеры, приложения стриминговых сервисов и др. Главным недостатком всех способов регулирования на стороне пользователя является не гибкость системы, пользователь все еще может получить доступ к деструктивному сайту, если тот отсутствует в черном списке.

Несмотря на разнообразие методов и технологий в области регулирования контента в интернете их недостатки мотивируют различные частные организации и государства продолжать разработки в этом направлении. На данный момент самой перспективной технологией как в регулировании контента, так и в других областях является искусственный интеллект.

Аналоги.

В таблице 1 перечислены некоторые популярные программы способные осуществлять регулирование доступа к деструктивным интернет ресурсам. В ниже приведенной таблице 1 используются следующие обозначения: «Нет» – функционал отсутствует; «Да» – функционал присутствует, параметр отражают характеристики на основании которых осуществлялось сравнение программных средств, наиболее важным в контексте данной работ является пункт характеризующие метод определения деструктивности контента. Черный список подразумевает то непосредственно сам анализ и выдача заключений касательно деструктивности контента является обязанностью человека.

Cold Turkey. Позволяет полностью или временно заблокировать доступ к сайтам, добавленным в соответствующий список. У приложения есть и платная версия, позволяющая блокировать также и приложения.

Родительский контроль Windows и macOS. Это встроенные в соответствующие операционные системы программы. Имеют схожий

принцип работы, а именно блокируют доступ к сайтам, которые пользователь внес в соответствующий список. Однако в случае Windows блокировка распространяется только на браузеры компании Microsoft. Parental Control Kroha — это приложение для родительского контроля, которое помогает управлять временем использования телефона, приложениями, социальными сетями и веб-сайтами.

Несмотря на широкий функционал, позволяющий производить достаточно индивидуальную настройку регулирования доступа к интернет-ресурсам, все они используют списки сайтов, которые необходимо вручную составлять пользователям, что хорошо работает с теми сайтами, которые уже уличены в распространении запрещенного контента. Однако в случае если сайт не был замечен ранее, то он не будет блокироваться. Частично эта проблема решается заранее составленным списками деструктивных сайтов, которые составляются другими пользователями или даже компаниями, такие списки как правило включают обширные перечни адресов, в которые входят как широко известные, так и наоборот, такие которые пользователь с легкостью может пропустить при составлении своего собственного списка, также данный подход помогает экономить время, и позволяет конкретному пользователю избежать взаимодействия с такими сайтами. Однако сохраняется необходимость периодически обновлять подобные списки, скачивая обновленные их версии из интернета или дорабатывая их вручную, кроме того подобные списки зачастую быстро теряют актуальность ввиду быстрого появления новых интернет-ресурсов, которые необходимо проверять вручную или же изначально не являются полными и содержат перечень наиболее известных сайтов с запрещенным контентом. Экспертная система, разработанная в рамках данной работы, имеет потенциальное преимущество над описанными выше программами, поставляемыми на гражданский рынок, т.к. при дальнейшем развитии может для своей

корректной работы не требовать от пользователя составления тех самых списков нежелательных ресурсов.

Таблица № 1

Сравнение аналогов разрабатываемой системы

Параметры	Cold Turkey, наличие	Родительский контроль Windows / macOS, наличие	Parental Control Kroha, наличие	Экспертная система, наличие
Назначение прав пользователям	нет	да	да	да
Метод регулирования контента: список ресурсов	да	да	да	нет
Метод регулирования контента: нейросети	нет	нет	нет	да
Бесплатная версия	да	да	нет	да
Платная версия/подписка	да	нет	да	нет
Анализ текста	нет	нет	нет	да
Анализ изображений	нет	нет	нет	нет
Анализ видео	нет	нет	нет	нет
Анализ аудио	нет	нет	нет	нет

Во всех программах пользователь, например, один из родителей ребенка должен самостоятельно производить отбор запрещенного контента или же искать подобные перечни в интернете, где аналогичную задачу выполняют другие пользователи, недостатки данного подхода были описаны выше, в то время, когда описываемая система может распознавать деструктивный контент без регулярной донастройки пользователем.

Главным преимуществом разработанной системы является в первую очередь доступность. Разработанная экспертная система, проверяет непосредственно те сайты, на которые пытается попасть пользователь, что занимает немного времени и может доставить некоторый дискомфорт, но сможет обеспечить защиту в том числе и от тех сайтов, которые еще не были заблокированы в Роскомнадзоре.

Разработка экспертной системы.

Общий алгоритм работы программного средства представлен на рис. 1.

На первом этапе пользователю предлагается ввести логин и пароль. В рассматриваемой версии программного средства пользователи делятся на администраторов и простых пользователей. Администраторы имеют расширенные права и для них не осуществляется фильтрация интернет-контента. Простые пользователи же имеют ограничение в правах и для них применяется фильтрация контента.

Мониторинг контента работает в фоновом режиме для простых пользователей по умолчанию, ожидая, когда пользователь попытается перейти на новый интернет-ресурс, алгоритм включает несколько этапов:

– Временная блокировка. Происходит сразу как пользователь пытается получить доступ к сайту и сохраняется до окончания проверки контента.

– Анализ контента. Важно отметить что в контексте данной работы и представленной в ней версии программы речь идет о текстовом контенте.

Начинается сразу после временной блокировки, по завершению контент классифицируется как безопасный и опасный.

– Блокировка. Если результат анализа покажет, что контент опасен, то доступ к сайту полностью блокируется, пользователь получает соответствующее уведомление.

– Разблокировка сайта. Происходит если сайт признается безопасным. Пользователю предоставляется доступ к сайту.

Описанный алгоритм мониторинга интернет ресурсов повторяется каждый раз, когда пользователь пытается получить доступ к какому-либо интернет-ресурсу, в том числе, если пользователь уже пытался успешно или безуспешно получить доступ к конкретному сайту [4].

Аутентификация пользователя реализована путем ввода пользователем логина и пароля. После ввода данных пользователя и нажатия кнопки вход в базу данных формируется запрос содержащий логин и хеш-функцию пароля. В базе данных пароль хранится в хешированном виде. Если из базы приходит пустой ответ – это говорит об отсутствии такого пользователя, иначе осуществляется вход. В настоящей версии приложения пользователи делятся на два типа: администраторы и простые пользователи (рис. 2).

У простых пользователей есть доступ к редактированию своего профиля – изменение логина и пароля, кроме того для данной категории пользователей в фоновом режиме осуществляется контроль доступа к деструктивным интернет-ресурсам. У администраторов ранее упомянутый контроль не осуществляется, кроме того у них есть доступ к функционалу, позволяющему осуществлять управление учетными записями пользователей: изменять пароль, логин и статус других пользователей или создать нового. Кроме того, можно создать нового пользователя из окна входа. Общий алгоритм аутентификации пользователей продемонстрирован на рис. 3.

В данной системе акцент делается на текстовом контенте графические и аудио материалы, размещенные на сайте программным средством, игнорируются. Непосредственно анализ текста осуществляется при помощи нейронных сетей. Алгоритм продемонстрирован на рис. 4.

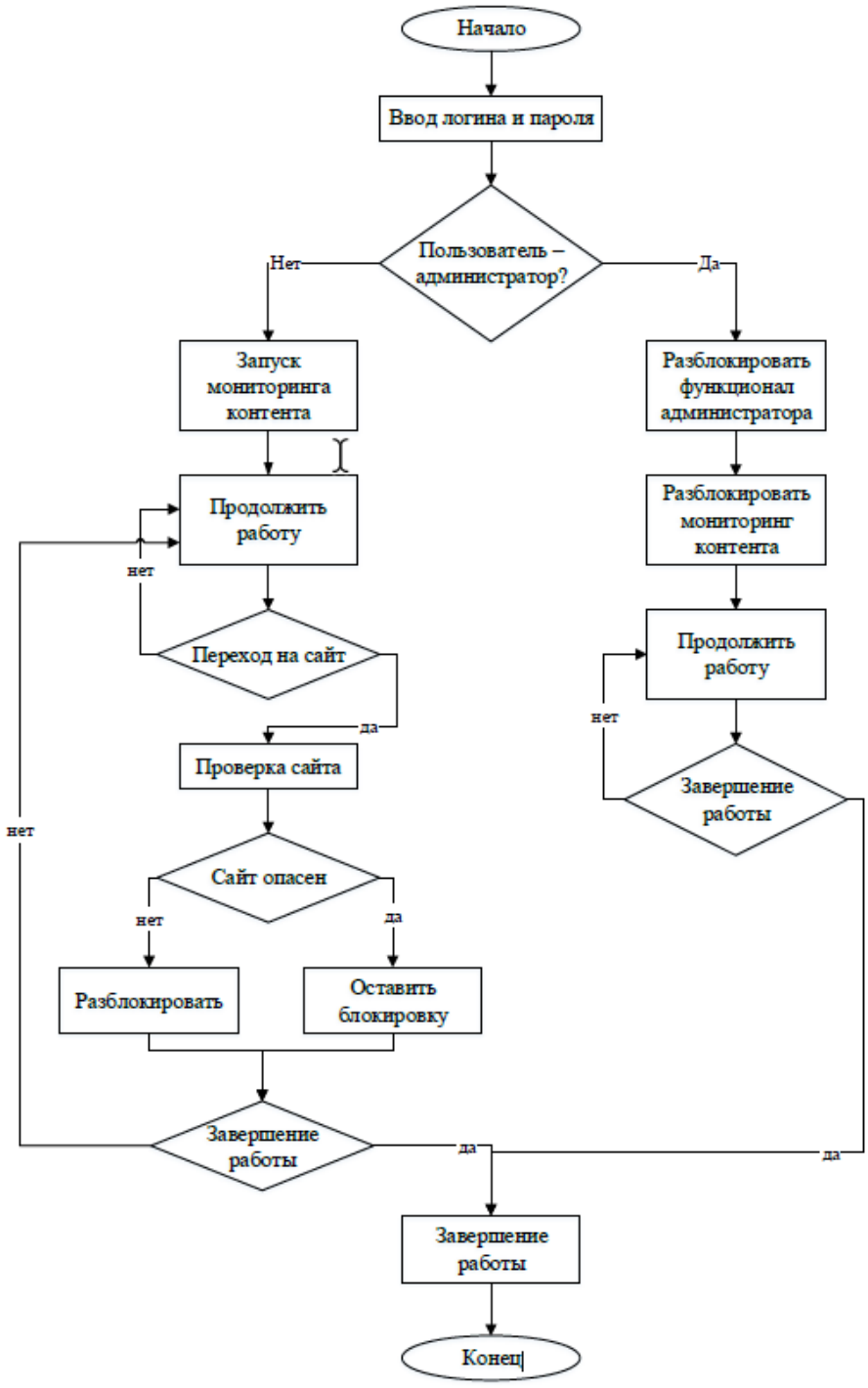

Рис. 1. – Общий алгоритм работы программы

Пользователи
id
login
password
status

Типы пользователей	
0	Админ
1	Пользователь

Рис. 2. – Данные пользователей

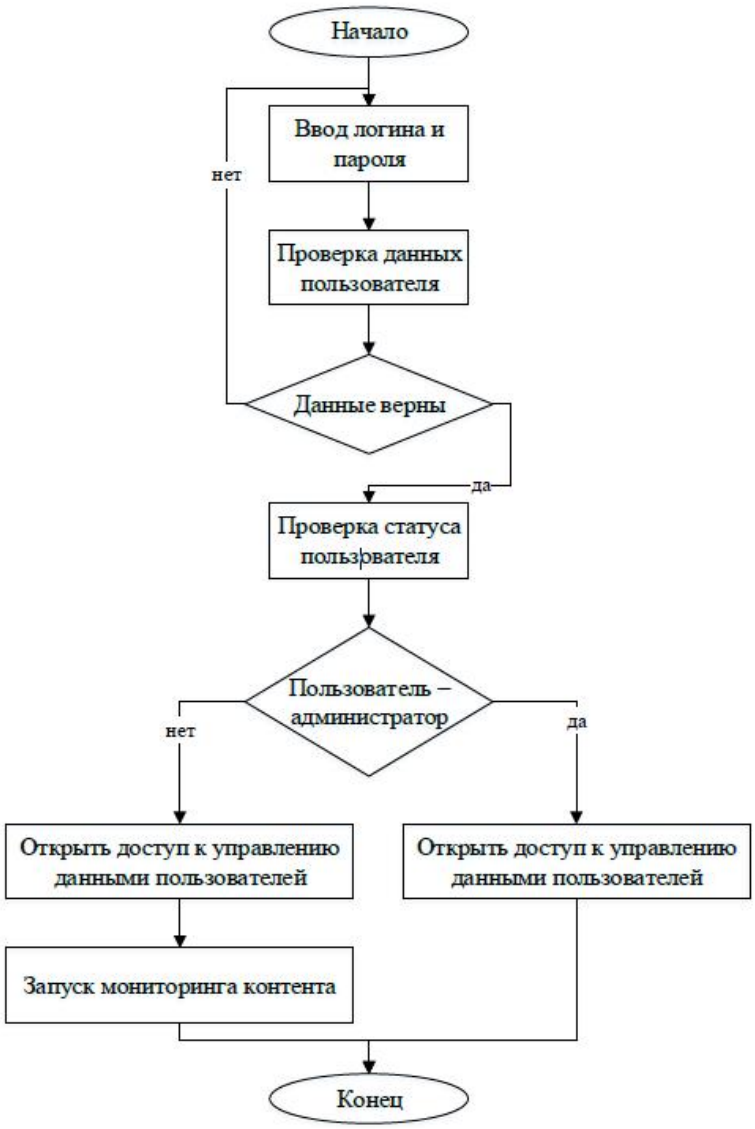

Рис. 3. – Алгоритм аутентификации пользователя

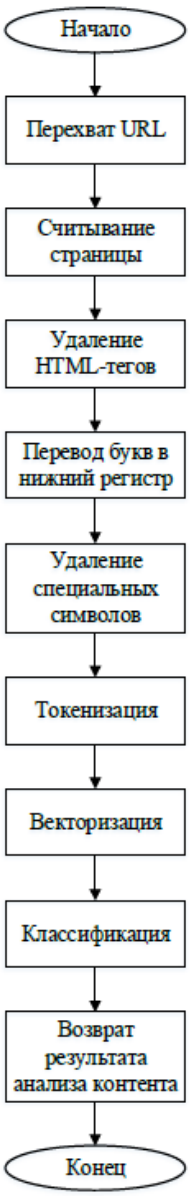

Рис. 4. – Алгоритм анализа текстового контента

Первый этап – это перехват URL и считывание страницы сайта для дальнейшей обработки контента.

Второй этап – предварительная обработка текста. Данный этап включает в себя следующие действия:

– Удаление html-тегов. Теги содержат множество информации о визуальном дизайне сайта, ссылки на сторонние интернет-ресурсы, ссылки

на изображения, видео и аудио, скрипты. Полезными для анализа могут быть только изображения, видео и аудио файлы, однако в рамках данной работы они не рассматриваются.

– Перевод всех букв в нижний регистр. Данная операция необходима, т.к. в понимании компьютера символы верхнего и нижнего регистра – это разные символы, как следствие, одно и тоже слово, написанное в двух разных вариантах будет считаться за разные слова, что приведет как к увеличению модели векторизатора, так и к уменьшению точности последующей классификации.

– Удаление специальных символов и знаков препинания [5]. В большинстве своем они не несут достаточное количество информации о тексте в контексте рассматриваемой задачи, кроме того, они снова могут привести к увеличению модели векторизатора и уменьшению точности последующей классификации.

– Токенизация. В рассматриваемой версии ПО текст разбивается на токены в размере одного слова [6]. Однако при дальнейшем изучении данного вопроса возможно использование и других размеров токенов.

– Лемматизация. Данный метод, в отличие от стемминга, исключает некоторые возможные ошибки, которые могут возникнуть в результате грубого отбрасывания частей слов, недостатком является повышенное время работы алгоритма, однако вычислительные мощности современных устройств делают данный недостаток малозаметным [7]. Второй недостаток – удаление формы слова удаляет и часть информации о контексте, в котором оно употреблено. Использование данного метода оправдывается относительно небольшим набором данных для обучения (датасетом), однако при использовании большей выборки данных возможен отказ от подобного метода.

Векторизация текста. Это обязательный этап необходимый для дальнейшей подачи текста на вход нейронной сети – классификатора.

Последний этап – непосредственно классификация текста для которой была использована нейронная сеть.

Для обучения нейронной сети необходимо подготовить датасет, который будет содержать образцы текста и их характеристику, представляющую собой числовой параметр отражающий категорию к которой данный конкретный образец относиться, в данной работе рассматриваются две категории: деструктивный контент и безопасный контент.

После загрузки обучающего датасета в программу над всеми образцами текста проводится предварительная обработка.

Далее проводится векторизация всех образцов и подача их на вход нейронной сети для обучения. Модели векторизации и классификации заранее либо проектируются и реализуются с нуля, либо используются уже реализованные алгоритмы, которые настраиваются для выполнения конкретной задачи. Использованные в данной работе алгоритмы будут описаны в последующих разделах.

Последним этапом является сохранение моделей векторизатора и классификатора для дальнейшего их использования в работе основной программы.

Общий алгоритм обучения нейронной сети для решения задачи классификации текстового контента, демонстрирующий все основные этапы данного процесса продемонстрирован на рис. 5.

Для обучения нейронной сети-классификатора использовался датасет состоящий из более 2000 образцов текста относящиеся к деструктивным или безопасным. На вход нейронной сети он подается в файле формата csv

содержащем два столбца образцы и метки, где 0 – метка, обозначающая безопасный материал, 1 – запрещенный.

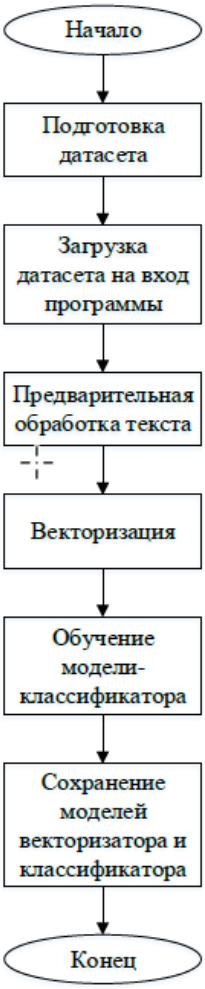

Рис. 5. – Алгоритм обучения нейронной сети-классификатора

Загрузка содержимого датасета в программу осуществляется при помощи функционала библиотеки «csv», специализирующейся на обработке файлов одноименного формата.

Предварительная обработка текста осуществляется при помощи следующих библиотек:

– «re» реализующий функционал, позволяющий работать с регулярными выражениями, в описываемом программном средстве используется для удаления html-тегов и специальных символов;

– «SpaCy» реализует различные методы для решения задач NLP, в данной работе используется для лемматизации слов [8];

– Векторизация текста реализована при помощи библиотеки Keras [9], модуль Tokenizer, помимо разбиения текста на токены (в данной работе за один токен принимается одно слово), также реализует алгоритм мешка слов для преобразования входного текста в числовую последовательность [10].

Для построения нейронной сети и ее обучения использован функционал библиотеки «Keras», предназначенной для глубокого машинного обучения.

Для анализа естественного языка хорошо подходят рекуррентные нейронные сети (RNN), а именно они могут обрабатывать тексты различной длины, так как, ввиду особенностей архитектуры, их размер не фиксирован. Это позволяет эффективно работать с текстами разной длины, от коротких предложений до длинных текстов. Кроме того, они хорошо справляются с последовательностями данных, к которым относятся и все естественные языки.

Однако RNN имеют свои ограничения, такие, как проблема затухания градиентов и ограничения в обработке долгих зависимостей.

LSTM (Долгая краткосрочная память) - это специальный тип рекуррентной нейронной сети, который обладает рядом преимуществ, делающих его популярным выбором для обработки последовательных данных в сравнении с обычными рекуррентными нейронными сетями.

Датасет, который был использован для обучения нейронной сети, состоит из 4000 комментариев пользователей социальных сетей, половина из которых содержит высказывания, являющиеся деструктивными, в то время как вторая половина безопасна. В ходе обучения был достигнут показатель 93,75% точности, что продемонстрировано на рис. 6, скриншоте проверки обучения.

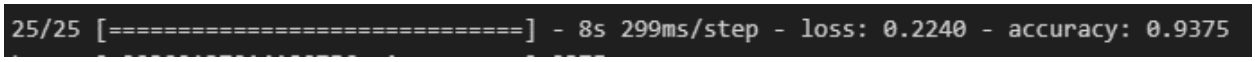

Рис. 6. – Точность обучения нейронной сети

Последний этап – сохранение моделей векторизации и классификации. Для этого был использован функционал Keras – save и load_model, позволяющий сериализовать или десериализовать объекты.

Сериализация – это процесс преобразования объекта в поток байтов, который затем может быть сохранен в файл или передан через сеть.

Десериализация – это обратный процесс, при котором поток байтов преобразуется обратно в объект.

В описываемой версии программы отсутствует функционал, который позволил бы переобучить имеющиеся модели. Поэтому единственный способ это сделать – обучить установить их вручную.

Окно входа предлагает пользователю ввести свои логин и пароль, для определения доступного ему функционала, в зависимости от его статуса в системе.

При неудачной попытке входа пользователь будет уведомлен об этом и сможет попробовать ввести свои учетные данные еще раз или создать новую учетную запись на основе тех данных, которые введены в соответствующие поля. Новая учетная запись по умолчанию будет иметь статус простого пользователя, т.е. для него будет осуществляться регулирование доступа к сайтам, у него не будет доступа к функционалу администратора, например, панели управления другими пользователями. Альтернативный способ создать новый аккаунт или восстановить утраченный доступ к существующему – обратиться к администратору, который может выдать новый пароль пользователю. Окно входа продемонстрировано на рис. 7.

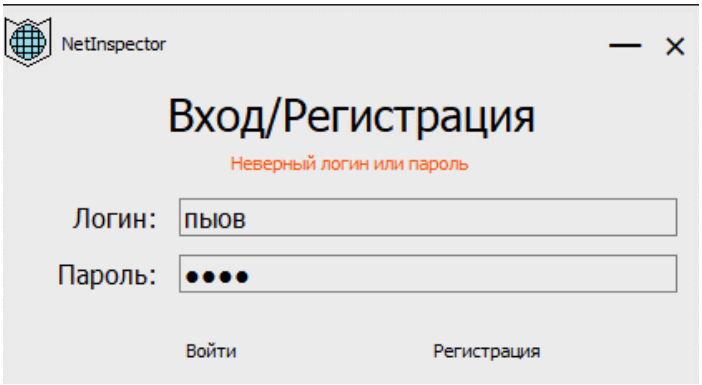

Рис. 7. – Окно входа

Для реализации пользовательского интерфейса использовалась библиотека PyQt5 [11]. Поле пароля автоматически скрывает данные пользователя.

Данные пользователя хранятся в локальной базе данных программного средства, работа с которой осуществляется при помощи библиотеки SQLite [12]. Пароль пользователя в целях безопасности хранится в виде хеш-функции, для чего использовалась библиотека HashLib. Т.к. такое кодирование не предполагает возможность обратной расшифровки в исходный текст, операция сравнения паролей осуществляется при помощи сравнения хеш-функций, кроме того, сохранение нового пароля осуществляется по такому же принципу, сохраняется не сам новый пароль, его закодированная версия.

По сути своей сообщения о блокировках – это простые html-страницы, которые будут открыты пользователю в случае блокировки сайта и сообщающие о причинах блокировки. Всего таких страниц две: предупреждение о временной блокировке и о постоянной.

Страница с сообщением о временной блокировке (рис. 8) показывается пользователю, когда он пытается перейти на новый сайт и сообщает ему о том, что страница проверяется на наличие деструктивного контента.

Страница с сообщением о постоянной блокировке (рис. 9) показывается если сайт, на который пытался зайти пользователь, признается деструктивным.

Данные страницы не содержат каких-либо ссылок на сторонние ресурсы или скрипты, обеспечивающие работу каких-либо функций программного средства, и являются по сути просто уведомлениями.

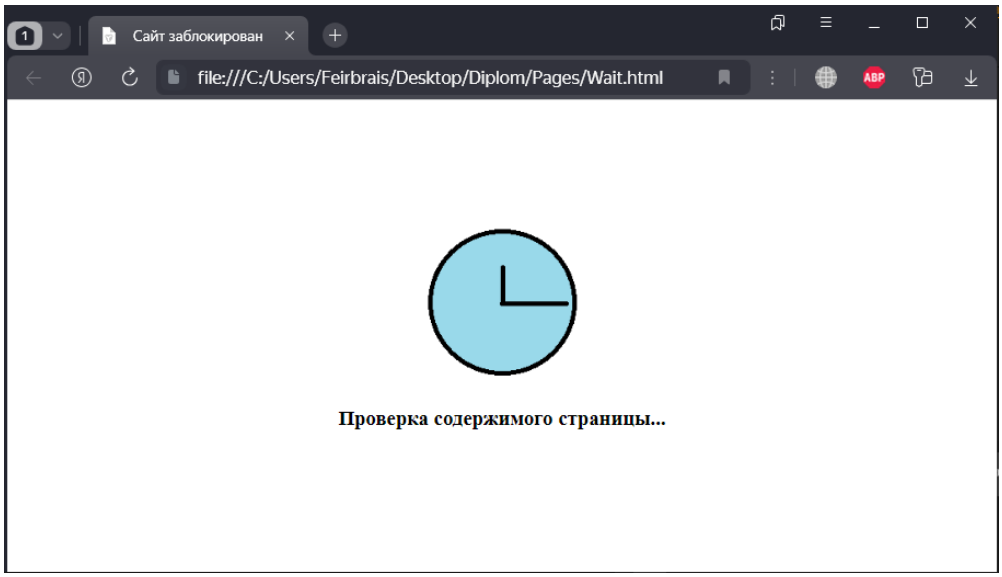

Рис. 8. – Сообщение о временной блокировке

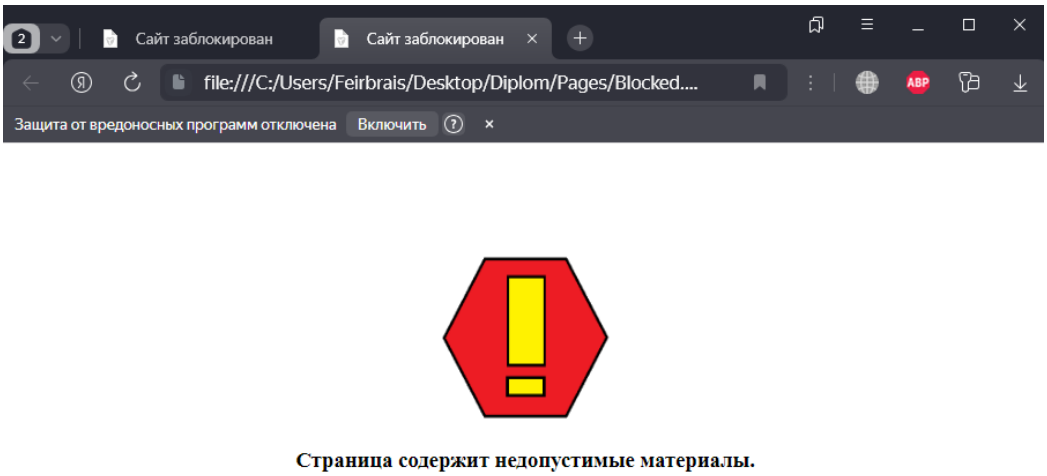

Рис. 9. – Сообщение о постоянной блокировке

Заключение.

Разработка экспертной системы регулирования доступа к деструктивным интернет-ресурсам является важным шагом в направлении повышения безопасности и благополучия пользователей интернета. Такая система позволит эффективно выявлять и блокировать доступ к ресурсам, содержащим вредоносный или запрещенный контент, тем самым снижая риск негативного воздействия на пользователей, особенно детей и подростков.

Важными преимуществами экспертной системы являются ее способность анализировать большие объемы данных, выявлять закономерности и принимать решения на основе сложных алгоритмов. Это позволяет ей оперативно реагировать на появление новых деструктивных ресурсов и адаптироваться к меняющимся условиям интернет-среды.

Однако, важно отметить, что эффективность экспертной системы зависит от качества ее разработки, точности алгоритмов и полноты базы данных. Поэтому, необходимо постоянно совершенствовать и обновлять систему, чтобы она могла успешно противостоять новым вызовам и угрозам.

В целом, экспертная система регулирования доступа к деструктивным интернет-ресурсам является важным инструментом в борьбе за безопасность и благополучие пользователей интернета. Ее разработка и внедрение будут способствовать созданию более безопасной и здоровой интернет-среды для всех.

Литература

1. Kemp S. Digital 2024: The Russian Federation // Datareportal. URL: datareportal.com/reports/digital-2024-russian-federation?rq=russia%20 (дата обращения 11.08.2024).

2. Крылова Е. Блок и сфера: объем контента с пропагандой насилия в интернете вырос вдвое // Известия URL: iz.ru/1632799/elizaveta–krylova/blok-

i-sfera-obem-kontenta-s-propagandoi-nasiliia-v-internete-vyros-vdvoe. (дата обращения 10.08.2024).

3. Jhaver S., Berman I., Gilbert E., Bruckman A. Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator // ACM Transactions on Computer-Human Interaction. 2019. Vol. 26, No. 5. URL: dl.acm.org/doi/fullHtml/10.1145/3338243.

4. Сердечный М.С. Разработка экспертной системы регулирования доступа к деструктивным интернет ресурсам // Студенческий вестник. – 2023. № 47-11 (286). С. 10–14.

5. Челышев Э. А., Оцоков Ш. А., Раскатова М. В., Щёголев П. Сравнение методов классификации русскоязычных новостных текстов с использованием алгоритмов машинного обучения // Вестник кибернетики. 2022. №1 (45). С. 63-71. URL: vestcyber.ru/jour/article/view/417.

6. Марков А.К., Семёночкин Д.О., Кравец А.Г., Яновский Т.А. Сравнительный анализ применяемых технологий обработки естественного языка для улучшения качества классификации цифровых документов // International Journal of Open Information Technologies. 2024. №3. С. 66-77. URL: injoit.org/index.php/j1/article/view/1769.

7. Khyani D., Siddhartha B.S. An Interpretation of Lemmatization and Stemming in Natural Language Processing // Journal of University of Shanghai for Science and Technology. 2021. № 22(10). pp. 350-357. URL: jusst.org/issue10-2020/.

8. Doc.spaCy API Documentation // spaCy.io URL: spacy.io/api/doc (дата обращения: 01.08.2024).

9. Keras 3 API documentation // Keras.io URL: keras.io/api/ (дата обращения: 01.08.2024).

10. Сердечный, М.С. Выбор алгоритма векторизации текста для решения задачи классификации // Актуальные проблемы науки и техники. 2023. С. 388–389.

11. Qt for Python Documentation // Qt.io. URL: doc.qt.io/-qtforpython-5/contents.html (дата обращения: 05.08.2024).

12. SQLite Documentation // SQLite.org URL: sqlite.org/-docs.html (дата обращения: 04.08.2024).

References

1. Kemp S. Digital 2024: The Russian Federation. Datareportal. URL: datareportal.com/reports/digital-2024-russian-federation?rq=russia%20 (accessed: 08/11/2024).

2. Kry`lova E. Blok i sfera: ob``em kontenta s propagandoj nasiliya v internete vy`ros vdvoe [Block and sphere: the volume of content with propaganda of violence on the Internet has doubled]. Izvestiya URL: iz.ru/1632799/elizaveta–krylova/blok-i-sfera-obem-kontenta-s-propagandoi-nasiliia-v-internete-vyros-vdvoe. (accessed: 08/10/2024).

3. Jhaver S., Berman I., Gilbert E., Bruckman A. ACM Transactions on Computer–Human Interaction. 2019. Vol. 26, No. 5. URL: dl.acm.org/doi/fullHtml /10.1145/3338243.

4. Serdechny`j M.S. Studencheskij vestnik - 2023. № 47-11 (286). pp. 10-14.

5. Chely`shev E`. A., Oczokov Sh. A., Raskatova M. V., Shhyogolev P. Vestnik kibernetiki. 2022. №1 (45). pp. 63–71. URL:vestcyber.ru/jour/article/view /417.

6. Markov A.K., Semyonochkin D.O., Kravecz A.G., Yanovskij T.A. International Journal of Open Information Technologies. 2024. №3. pp. 66-77 URL: injoit.org/index.php/j1/article/view/1769.

7. Khyani D., Siddhartha B.S. Journal of University of Shanghai for Science and Technology. 2021. № 22(10). pp. 350-357. URL: jusst.org/issue10-2020/.

8. SpaCy.io URL: spacy.io/api/doc (accessed: 08/01/2024).

9. Keras.io URL: keras.io/api/ (accessed: 08/01/2024).

10. Serdechny`j, M.S. Aktual`ny`e problemy` nauki i texniki. 2023. pp. 388–389.

11. Qt.io. URL: doc.qt.io/–qtforpython–5/contents.html (accessed: 08/05/2024).

12. SQLite.org. URL: sqlite.org/-docs.html (accessed: 08/04/2024).

Дата поступления: 14.07.2024

Дата публикации: 25.08.2024