

АНАЛИЗ ОСНОВНЫХ МЕТОДОВ ВЕБ-ФИЛЬТРАЦИИ КОНТЕНТА НА ПРИМЕРЕ DETOX BROWZER И АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

С.С. Мельниченко, Московский технический университет связи и информатики, sere2300@gmail.com.

УДК 004.8

Аннотация. Ежегодно потребление информации из различных веб-ресурсов растет очень быстрыми темпами, главным образом, из-за повышения доступности и распространения интернета. Многие люди полагаются на интернет, чтобы получить различные точки зрения на интересующие их вопросы, но в то же время они могут быть подвергнуты воздействию контента, который негативно повлияет на их психическое благополучие. Броские заголовки и эмоционально заряженные статьи увеличивают число читателей, что, в свою очередь, увеличивает доход от рекламы для веб-сайтов. Когда пользователь потребляет большое количество негативного контента, это отрицательно влияет на его жизнь в целом и оказывает значительное влияние на его настроение и душевное состояние. Многие исследования, проведенные во время пандемии *COVID-19*, показали, что люди по всему миру, независимо от страны их происхождения, испытывали более высокий уровень тревоги и депрессии. Методы фильтрации контента могут помочь в создании цифровой среды, которая больше подходит для людей, склонных к депрессии, беспокойству и стрессу. Значительный объем работы был проделан в области веб-фильтрации, но уделялось ограниченное внимание оказанию помощи особо чувствительным лицам. В данной статье рассмотрен *Detox Browser* – простой инструмент, который позволяет конечным пользователям контролировать свое воздействие на темы, которые могут повлиять на их психическое благополучие. Помимо этого, в работе рассмотрены основные методы фильтрации контента в веб-пространстве.

Ключевые слова: веб-фильтрация; контент; нейронные сети; машинное обучение; *Detox Browser*; фильтрация по ключевым словам.

ANALYSIS OF THE MAIN METHODS OF WEB CONTENT FILTERING ON THE EXAMPLE OF DETOX BROWZER AND MACHINE LEARNING ALGORITHMS

Sergey Melnichenko, Moscow technical university of communications and informatics.

Annotation. Every year, the consumption of information from various web resources is growing very rapidly, mainly due to the increased availability and spread of the Internet. Many people rely on the internet to get different points of view on issues of interest to them, but at the same time they may be exposed to content that may be harmful to their mental well-being. Catchy headlines and emotional articles increase the number of readers, which, in turn, increases advertising revenue for websites. When a user consumes a large amount of negative content, it negatively affects his life in general and has a significant impact on his mood and state of mind. Many studies conducted during the *COVID-19* pandemic have shown that people around the world, regardless of their country of origin, experienced higher levels of anxiety and depression. Content filtering methods can help in creating a digital environment that is more suitable for people prone to depression, anxiety and stress. A significant amount of work has been done in the field of web filtering, but limited attention has been paid to helping particularly sensitive individuals. This article considers *Detox Browser*, a simple tool that allows end users to

monitor their exposure to topics that may affect their mental well-being. In addition, the paper discusses the main methods of filtering content in the web space.

Keywords: web filtering; content; neural networks; machine learning; Detox Browser; keyword filtering.

Введение

Чтобы решить проблемы, возникающие в результате длительного воздействия негативного контента, был концептуализирован и разработан *Detox Browser*, позволяющий пользователю отфильтровывать контент, который, по его мнению, негативно влияет на его психическое благополучие. В дополнение к фильтрам по умолчанию, расширение позволяет пользователю настраивать просмотр в соответствии со своими личными требованиями. Расширение анализирует настраиваемое, классифицирует контент и удаляет темы, внесенные в черный список, чтобы пользователь не был перегружен информацией, которая может его травмировать. Расширение производит фильтрацию с помощью простых проверок на основе ключевых слов с последующим подробным анализом. Несмотря на то, что он в основном нацелен на результаты поиска *Google*, он также проверяет веб-сайты на наличие тем, внесенных пользователем в черный список. Расширение находится в открытой бета-версии, и его функции постоянно совершенствуются. Помимо этого, в статье рассмотрены методы машинного обучения, применяемые при анализе контента с целью его дальнейшей фильтрации.

Проектирование и разработка

Detox Browser был создан как расширение для браузера *Chrome*, предназначенное, в первую очередь, для фильтрации результатов поиска *Google*. Он также поддерживает обнаружение ненормативной лексики и предупреждения о содержимом для межсайтовой навигации. Поведение расширения в каждом контексте можно изменить с помощью настроек, а чувствительность можно настроить с помощью параметров персонализации. Как только *Google*-поиск обнаруживает что-либо на основе *URL*-адреса, расширение извлекает все *HTML*-узлы со страницы, на которые есть ссылки. Эти извлеченные ссылки затем используются для получения ближайших родительских узлов. Они классифицируются в зависимости от того, какому из predefined шаблонов они соответствуют. Эти шаблоны для категоризации получаются путем ручного анализа селекторов для требуемых компонентов анализируемой веб-страницы. В настоящее время эти шаблоны были извлечены для *Google*-поиска, и расширение поддерживает обычные результаты поиска, избранные истории, новости и видео, игнорируя при этом специальные элементы и результаты, например, википедии или словаря, так что прямой поиск по теме и ее значению не блокируется.

Наблюдатель за изменениями, представляющий собой встроенный объект, который отслеживает любые изменения в элементе объектной модели документа (*Document object model, DOM*), гарантирует, что любые изменения в содержимом страницы поиска проверяются расширением. Из родительских узлов текстовое содержимое извлекается путем рекурсивного обхода *DOM*. Затем этот текст проверяется с помощью анализа настроений на основе ключевых слов, основанного на *AFINN*. *AFINN* – это самый простой, но популярный словарь, используемый для анализа настроений, разработанный Финном Арупом Нильсеном. Он содержит более 3300 слов с оценкой полярности, связанной с каждым словом. В языке программирования *Python* есть встроенная функция для этого словаря. Анализ настроений может быть выполнен либо с помощью методов машинного обучения, либо с помощью помеченного набора данных, в случае рассматриваемого

расширения используется последнее для сохранения малого размера расширения и скорости работы. Некоторые из популярных инструментов, основанных на лексиконе, включают *VADER*, *SentiWordNet* и *AFINN*. В расширении *Detox Browser* был использован *AFINN*, поскольку он хорошо работает даже при отрицательных оценках выборки. Набор данных *AFINN* содержит слова с рейтингом от -5 (отрицательный) до +5 (положительный), что позволяет проводить быстрый предварительный анализ, чтобы облегчить пометку статей при загрузке страницы на основе того, окажется ли значение отрицательным. Помеченные статьи заменяются *div*-заполнителем, по которому можно щелкнуть, чтобы восстановить замененный элемент. Эмодзи, изображающие сильно негативный, негативный, нейтральный, позитивный и сильно позитивный, добавляются слева от статей, указывающих на оценку, полученную с помощью лексического анализа.

Для дальнейших проверок извлеченный текст передается через многочленный наивный байесовский классификатор и процессор естественного языка. Чтобы сохранить компактность расширения и упростить их обновление и улучшение, они развертываются онлайн и обслуживаются через конечную точку программного интерфейса приложения (*Application program interface, API*). Классификатор обучен на основе 20-летнего набора данных *Times of India Headlines* и классифицирует статьи по 50 лучшим категориям из набора данных, который насчитывает более 300 групп. Пользователь может навести курсор на содержимое, скрытое расширением, чтобы увидеть ключевые слова, сгенерированные с помощью подходов нейролингвистического программирования (НЛП). В заполнителе также упоминается доменное имя исходной статьи, чтобы помочь пользователю определить, хочет ли он увидеть статью.

Всплывающее окно расширения позволяет пользователям настраивать чувствительность по своему вкусу. Список полярности позволяет указывать фразы вместе со значением от -5 до 5, которые используются для переопределения значений по умолчанию из *AFINN*. Кроме того, черный список позволяет пользователям полностью удалять темы, которые они, возможно, не хотели бы видеть. В зависимости от параметров, выбранных в меню, поведение черного списка меняется. Слова или фразы в черном списке ищутся с помощью запросов регулярных выражений, и при обнаружении они по умолчанию размываются с возможностью удаления при наведении курсора. Для результатов поиска, если размытие отключено, расширение полностью удалит результат, содержащий ключевые слова, внесенные в черный список. Такой способ также используется для фильтрации ненормативной лексики и оскорбительных выражений. Для этого используется пакет ненормативных слов, который содержит большую коллекцию ненормативной лексики на разных языках. Расширение также содержит фоновый скрипт, который предупреждает пользователя, если какое-либо из внесенных в черный список ключевых слов обнаружено на веб-сайте, отличном от *Google*-поиска, который пользователь посещает с помощью всплывающего окна. Через панель настроек можно отключить расширение на определенных веб-сайтах.

Методы машинного обучения для фильтрации контента

В табл. 1 представлены алгоритмы машинного обучения, применяемые при фильтрации контента.

Таблица 1.

Контролируемое обучение	Неконтролируемое обучение	Усиленное обучение
Обучение производится на примерах верных	На вход подаются данные. Обучение происходит на основе поиска закономерностей.	Обучение происходит посредством системы

Контролируемое обучение	Неконтролируемое обучение	Усиленное обучение
ответов на аналогичной задаче вопросы.		положительных и отрицательных оценок работы.

Контролируемое обучение в настоящее время является наиболее популярным подходом. При таком подходе система обучается посредством «наблюдения». Ей дается обучающий набор, то есть большой набор (вероятно) правильных ответов на задачу системы, и она учится отвечать на новые случаи аналогичным образом. Например, обучающий набор для системы, предназначенной для обнаружения рекламного спама, состоит из сообщений, которые люди классифицировали как экземпляры спама. Система обнаружения спама учится классифицировать новые сообщения таким же образом, как и в своем обучающем наборе. Аналогичным образом, система, предназначенная для обнаружения разжигания ненависти (или порнографии), может быть обучена через набор сообщений, где случаи ненависти (или порнографии) отличаются от остальных сообщений.

При неконтролируемом обучении системы искусственного интеллекта (ИИ) обучаются, не получая внешних инструкций: вместо этого они выявляют закономерности в данных. Методы неконтролируемого обучения используются, в частности, для кластеризации, то есть для группировки набора элементов, которые представляют соответствующие сходства или связи (например, документы, относящиеся к одной и той же теме, люди, разделяющие соответствующие характеристики, или термины, играющие одинаковые концептуальные роли в текстах). Например, документы, содержащие один и тот же оскорбительный язык, могут быть автоматически сгруппированы или реклама, использующая заголовки, содержащие определенные слова.

При усиленном обучении система извлекает уроки из результатов своих собственных действий, а именно, она наблюдает за результатами своих действий и самостоятельно управляет вознаграждениями или штрафами (например, набранными или потерянными очками), которые связаны с результатами таких действий.

В машинном обучении применяется множество методов: деревья решений, статистическая регрессия, машина опорных векторов, эволюционные алгоритмы, методы усиленного обучения и так далее. В последнее время глубокое обучение, основанное на многоуровневых нейронных сетях, очень успешно внедряется, особенно там, где необходимо распознавать шаблоны и привязывать их к классификациям и решениям, как в случае фильтрации (например, классифицировать контент как оскорбительный или порнографический и относиться к нему соответствующим образом, то же самое возможно применять в отношении рекламных баннеров). Нейронные сети состоят из набора узлов, называемых нейронами, расположенных в несколько слоев и соединенных ссылками. Они называются так потому, что воспроизводят некоторые аспекты нервной системы человека, которая действительно состоит из взаимосвязанных специализированных клеток, биологических нейронов, которые получают и передают информацию. Каждый нейрон получает сигналы от подключенных нейронов или извне. В случае фильтрации входные данные могут представлять точки на изображении или слова в тексте. Каждый нейрон применяет некоторые вычисления к входным данным, которые он получает и, если результат достигает порогового значения нейрона, нейрон активируется, посылая сигналы подключенным нейронам или за пределы сети. Активация начинается с узлов, получающих внешние входные данные, и распространяется по сети. Обучение сети

происходит путем информирования сети о том, являются ли ее ответы (ее выходные данные) правильными или неправильными. Если ответ сети неверен, алгоритм обучения обновляет сеть, то есть корректирует веса связей между нейронами, чтобы в следующий раз, когда сети будут представлены эти входные данные, она выдала правильный ответ.

Существует множество нейронных сетей. В последние годы структура сетей становится все более сложной: были разработаны сети, которые имеют много уровней (так называемые глубокие нейронные сети), или которые учитывают близость входных данных, обычно точек на изображениях и объектов, возникающих из них (сверточные сети), или которые имеют возможность хранить и обрабатывать последовательности входных данных с учетом их порядка (рекуррентные нейронные сети или сети долговременной и кратковременной памяти).

Нейронные сети обладают наилучшей производительностью в задачах распознавания образов, таких как те, которые связаны с фильтрацией. К сожалению, они не дают объяснений своих результатов. Можно выяснить, как определенная информация была получена в результате активации сети, но эта информация не дает обоснования, значимого для людей: она не говорит нам, почему был дан определенный ответ. Таким образом, нейронные сети аналогичны черным ящикам и их применение далеко не всегда уместно в фильтрации контента, в том числе и поэтому. На рис. 1 показано упрощенное представление многослойной нейронной сети (в реальных сетях может быть гораздо больше слоев нейронов).

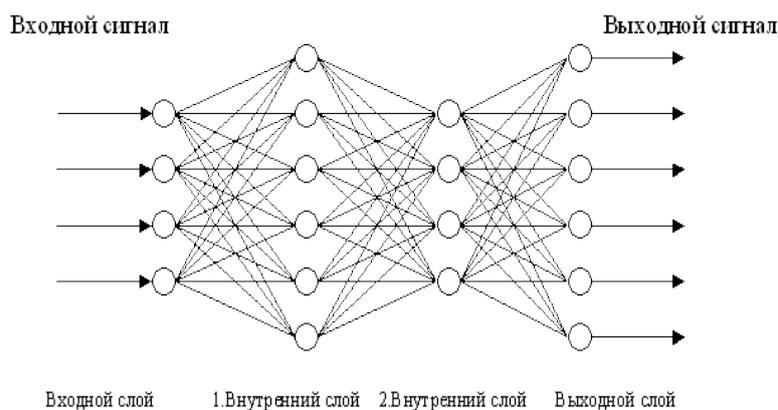


Рисунок 1

Заключение

В данной статье описан принцип функционирования расширения *Detox Browser* для *Google Chrome*, которое фильтрует результаты поиска в соответствии с предпочтениями пользователя. Оно также выдает всплывающее предупреждение, если контент на каком-либо веб-сайте занесен пользователем в черный список. Добавление этих прямых скриптов обеспечивает гораздо большую гибкость управления контентом, доставляемым на страницу. Кроме того, поскольку классификатор и модель НЛП развернуты на стороне сервера, есть возможность постоянно обновлять классификатор и набор инструментов для повышения его точности. Также были рассмотрены основные алгоритмы машинного обучения, используемые при фильтрации контента.

Литература

1. Эртурул Акбас. Фильтрация следующего поколения: Автономная фильтрация улучшенная архитектура прокси-сервера для фильтрации веб-контента // 23-й

- Международный симпозиум по компьютерным и информационным наукам. IEEE, 1-4, 2008.
2. Ма Аш-Шаби. Оценка эффективности наиболее важных лексиконов, используемых для анализа настроений и анализа мнений. IJCSNS 20, 1 (2020), 1.
 3. Саманта К. Брукс, Ребекка К. Вебстер, Луиза Э. Смит, Лиза Вудленд, Саймон Вессели, Нил Гринберг и Гидеон Джеймс Рубин. Психологическое воздействие карантина и как его уменьшить: быстрый анализ фактических данных // The Lancet 395, 10227, 2020. – С. 912-920.
 4. Манмун Де Чоудхури, Мередит Рингел Моррис и Райен Уайт. Поиск и обмен медицинской информацией в интернете: сравнение поисковых систем и социальных сетей // В материалах конференции SIGCHI по человеческому фактору в вычислительных системах, 2014. – С. 1365-1376.
 5. Каролина Фигероа и Адриан Агилера. Необходимость технологической революции в области психического здоровья в условиях пандемии COVID-19 // Границы в психиатрии 11, 2020.
 6. Болябкин М.В. Разработка и внедрение общего анализатора SQL // Международный журнал гуманитарных и естественных наук, 2022. – № 1-1 (64). – С. 55-61. – DOI 10.24412/2500-1000-2022-1-1-55-61. – EDN PRDDVT.
 7. Болябкин М.В. Интеллектуальная система для преобразования запросов на естественном языке в SQL и их выполнения // Международный журнал гуманитарных и естественных наук, 2021. – № 12-1(63). – С. 134-138. – DOI 10.24412/2500-1000-2021-12-1-134-138. – EDN OZNWEG.
 8. Mohamad H. Hassoun. Fundamentals of Artificial Neural Networks. MIT Press, Cambridge, Massachusetts, 1995.
 9. Хайкин С. Нейронные сети: полный курс. 2-е изд. – М.: «Вильямс», 2006.
 10. Haykin S. Neural Networks and Learning Machines. 3rd Edition. Pearson, 2018.
 11. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применения в экономике и бизнесе. Электронная версия книги находится здесь. – М., МИФИ, 1998.
 12. Тархов Д.А. Нейронные сети. Модели и алгоритмы. – М., Радиотехника, 2005. (Научная серия «Нейрокомпьютеры и их применение», ред. А.И. Галушкин. Кн.18.)
 13. Васильев А.Н., Тархов Д.А. Нейростоевое моделирование. Принципы. Алгоритмы. Приложения. СПб.: Изд-во Политехн. Ун-та, 2009. ISBN 978-5-7422-2272-9.
 14. Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры. – М.: Изд-во МГТУ им. Баумана, 2004.
 15. Галушкин А.И. Нейронные сети. Основы теории. – М.: Горячая линия - Телеком, 2010.
 16. Головкин В.А. Нейронные сети: обучение, организация и применение. – М.: ИПРЖР, 2001.
 17. Яхьяева Г.Э. Основы теории нейронных сетей. Интернет-университет информационных технологий, изд-во «Открытые системы».
 18. Круглов В.В., Длин М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. Физматлит, 2001.
 19. Aggarwal C.C. Neural Networks and Deep Learning. A Textbook. Springer International Publishing.
 20. Шумский С.А. Машинный интеллект. Очерки по теории машинного обучения и искусственного интеллекта. – М., РИОР, 2019. DOI: 10.29039/02011-1.
 21. Isakov Yu. A. Artificial intelligence // Modern Science, 2018. – No 6-1. – P. 25-27. – EDN RVEVVF.
 22. Vadinský O. An Overview of Approaches Evaluating Intelligence of Artificial Systems // Acta Informatica Pragensia, 2018. – Vol. 7. – No 1. – P. 74-103. – DOI 10.18267/j.aip.115. – EDN XWDXKXJ.

23. Адлер Ю.П. Алгоритмически неразрешимые задачи и искусственный интеллект // Экономика и управление: проблемы, решения, 2018. – Т. 7. – № 5. – С. 17-24. – EDN XTUTED.
24. Богомолов А.И. Искусственный интеллект и экспертные системы в мобильной медицине // Хроноэкономика, 2018. – № 3 (11). – С. 17-28. – EDN XUXVFB.
25. Борисова Е.В. Современный тренд образовательной среды – искусственный интеллект и цифровая педагогика // Традиции и новации в профессиональной подготовке и деятельности педагога: сборник научных трудов Всероссийской научно-практической конференции, Тверь, 29-30 марта 2018 года. – Тверь: Тверской государственный университет, 2018. – С. 84-87. – EDN UUYLPA.
26. Гаджиева А.Г. Цифровизация и занятость: роль отраслей сектора услуг // Инновации, 2018. – № 2. – С. 61-70
27. Галлямова И.Р. Информационная поддержка принятия решений при управлении лояльностью потребителей с использованием технологий искусственного интеллекта // Проблемы современной науки и образования, 2018. – № 7 (127). – С. 25-27. – EDN RUJXAJ.
28. Демкин В.И. Искусственный интеллект в робототехнике // Вестник современных исследований, 2018. – № 6.3 (21). – С. 456-458. – EDN XURODJ.
29. Демкин В.И. История и перспективы развития нейронных сетей // Вестник современных исследований, 2018. – № 6.1 (21). – С. 366-368. – EDN XURJSX.