

2. Иванов К. К. Проектирование информационных систем / К. К. Иванов. — Текст : электронный // Молодой ученый. — 2017. — № 19 (153). — С. 22-24. — URL: <https://moluch.ru/archive/153/43309/>
3. Федоров И. Г. Моделирование бизнес-процессов в нотации BPMN 2.0 : научно-практическое издание / И. Г. Федоров ; Министерство образования и науки Российской Федерации, Московский государственный университет экономики, статистики и информатики (МЭСИ). — Москва : МЭСИ, 2013. — 264 с. — Текст : непосредственный.
4. Цуканова О. А. Методология и инструментарий моделирования бизнес-процессов : учебное пособие / О. А. Цуканова. — Санкт-Петербург : Университет ИТМО, 2015. — 101 с. — Текст : непосредственный.
5. Дронов В. А. PHP, MySQL, HTML5 и CSS 3. Разработка современных динамических Web-сайтов / В. А. Дронов. — Санкт-Петербург : БХВ-Петербург, 2016. — 677 с. — Текст : непосредственный.
6. Рогов Е. В. PostgreSQL изнутри / Е. В. Рогов. — Москва : ДМК Пресс, 2022. — 660 с. — Текст : непосредственный.
7. Стаффер М. Laravel. Полное руководство / М. Стаффер. — 2-е изд. Санкт-Петербург : Издательский дом «Питер», 2021. — 512 с. — Текст : непосредственный.

М. Д. Долгушин, Д. К. Исмакова, П. К. Моор

Тюменский государственный университет, г. Тюмень

УДК 004.89

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ЭКСПЕРТНОГО АНАЛИЗА СООБЩЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ

***Аннотация.** В работе представлено решение задачи по разработке информационной системы, объединяющей экспертный и автоматический анализ сообщений из социальной сети для создания и исследования новых наборов данных.*

***Ключевые слова:** информационная система, обработка естественных языков, экспертная оценка, классификация текстов.*

Введение. С увеличением количества информации в социальных сетях растут проблемы, связанные с оскорбительной средой, токсичными сообщениями [1], в том числе и в таких российских

системах, как ВКонтакте, а значит задача создания распределенной системы анализа сообщений из социальных сетей является актуальной.

На данный момент в сети Интернет присутствует множество [2] наборов данных на русском языке, которые могли бы быть использованы для одноклассовой классификации и многоклассовой классификации отдельных характеристик сообщений в социальных сетях, таких как эмоциональный окрас, токсичность, сексизм и иные. Актуального совмещенного набора данных с множеством размеченных классов, характеризующих токсичные сообщения, как сексистские, как оскорбительные для личности и как оскорбительные для группы с определенной идентичностью, на русском языке достаточных размеров (для каждого класса должно быть ~10000 сообщений) не имеется, хотя существуют аналогичные на английском [3]. Данный набор мог бы значительно улучшить качество классификации оскорбительных сообщений, что в свою очередь повлияло бы на общую благоприятность общения в социальных сетях.

Проблема исследования. Для создания подобного набора данных возможно использовать существующие системы разметки данных большим количеством людей, такие как: Яндекс.Толока, Amazon Mechanical Turk [4] — однако сильная субъективность отношения многих людей к проблематике токсичных сообщений может значительно повлиять на качество разметки.

Материалы и методы. В данной работе рассматривается создание информационной системы экспертного анализа сообщений в социальных сетях, объединяющей методы автоматической и экспертной оценки. Автоматическая оценка классификаторами, позволит быстро разметить наборы данных, но достаточно неточно, поэтому данные предлагается дополнительно проверять с помощью небольшого количества экспертов, что повысит суммарное качество набора данных.

На рис. 1 представлена структура разработанной информационной системы.

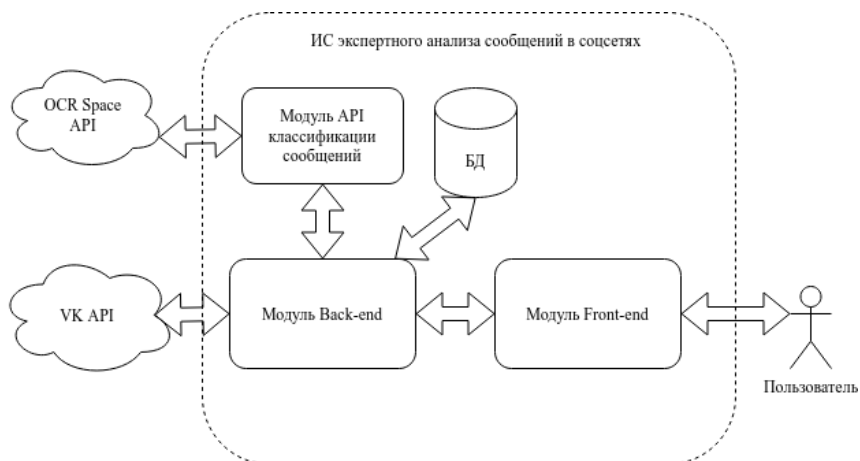


Рис. 1. Структура информационной системы экспертного анализа сообщений в социальных сетях

Здесь БД — база данных с таблицами по сообщениям, пользователям и оценкам пользователей на сообщения, реализованная в СУБД PostgreSQL. Взаимодействует с модулем Back-end.

В БД были выделены следующие сущности: Users — пользователи; Roles — роли у пользователей; Posts — посты; Comments — комментарии; Pictures — изображения; Post_Marks — оценки у поста; Comment_Marks — оценки у комментария; Pictures_Marks — оценки у изображения; Message_Topics — темы сообщений; Posts_Message_Topics — набора тем постов; Comments_Message_Topics — набора тем у комментариев; Pictures_Message_Topics — набор тем у изображений.

Модуль API классификации сообщений — API, разработанное на Python с фреймворком Flask. Очищает, векторизует и классифицирует сообщения, представленные изображениями, текстовыми сообщениями, что позволяет быстро разметить выборку любых размеров. Выполняет проверку на выбросы и очистку от них для набора данных по оценкам пользователей сообщениям. Взаимодействует со сторонним OCR Space API для распознавания текста на изображениях и модулем Back-end.

Модуль API классификации использует ансамбль из нескольких классификаторов методом градиентного бустинга CatBoost [5] для получения различных меток текстового сообщения, таких как: токсично, токсично-сексично, токсично для личности, токсично для группы, угроза. Также использовалась библиотека Dostoevsky для получения эмоционального окраса текста сообщения. В табл. 1 представлены результаты оценки точности обученных классификаторов.

Таблица 1

Точность классификации

Оценка	Токсично	Токсично-сексично	Токсично для личности	Токсично для группы	Угроза
Accuracy	0.778	0.9	0.939	0.868	0.978
Precision	0.783	0.896	0.935	0.867	0.977
Recall	0.778	0.9	0.939	0.868	0.978
F1	0.752	0.893	0.936	0.867	0.977

Результаты. Алгоритм классификации текстового сообщения представлен на рис. 2. Алгоритм классификации изображения соответствует ему за исключением того, что предварительно, на этапе добавления сообщения типа изображение в систему, с изображения считывается текст с помощью API Space OCR.

Далее был разработан модуль Back-end — REST API на языке программирования TypeScript с фреймворком Express, контролирующей работу пользователя с базой данных системы посредством ORM отображения в соответствии с правами роли пользователя. Также обеспечивает пересылку данных для обработки в модуле API классификации сообщений. Взаимодействует со сторонним VK API для получения дополнительной информации о сообщении из социальной сети ВКонтакте, с базой данных, с модулем API классификации сообщений, модулем Front-end.

Модуль Front-end — пользовательское приложение, реализованное с помощью библиотеки React на языке программирования TypeScript. Взаимодействует с модулем Back-end и пользователем

системы. Также данный модуль подключен к облачному хранилищу Firebase, используемому для аутентификации пользователя по протоколу OAuth, повышающему устойчивость системы к взлому.

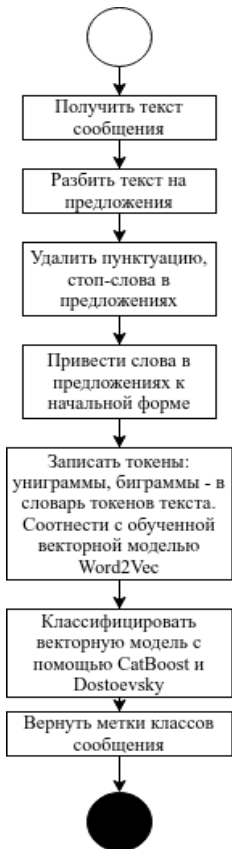


Рис. 2. Алгоритм классификации текста сообщения

На рис. 3 представлена страница разметки комментариев пользователя с ролью Эксперт.

На рис. 4 представлена страница сводной таблицы по постам для пользователя, зашедшего с ролью Исследователь, что обеспечивает ему такие функции как экспорт таблицы в CSV-файл, а также

возможность проверить качество набор данных и повысить качество набора данных, за счет удаления некорректных ответов экспертов из выборки.

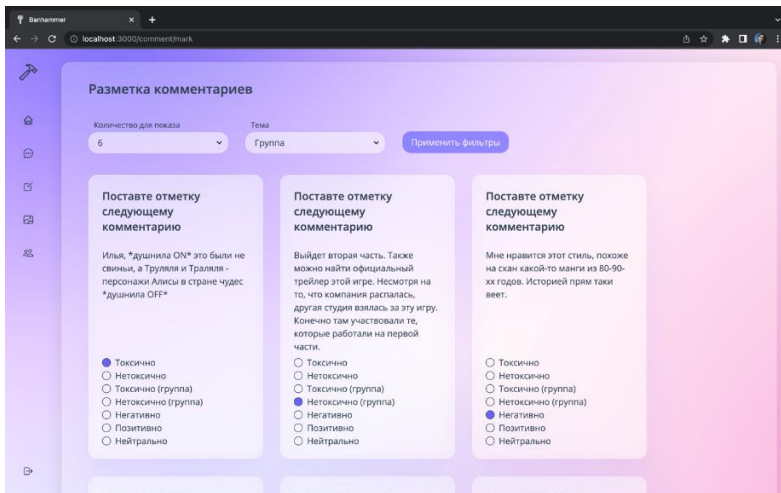


Рис. 3. Страница разметки комментариев

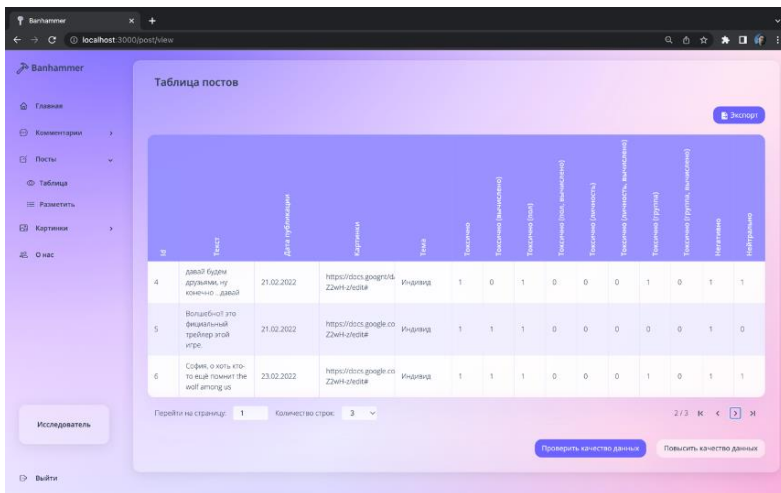


Рис. 4. Страница таблицы постов

Заключение. В данной статье представлена разработанная Web-система экспертного анализа сообщений в социальных сетях, объединяющая методы автоматической и экспертной оценки для быстрого получения наборов данных.

В процессе работы была создана ассамблея классификаторов методом CatBoost для разметки сообщений со следующими метками: токсично, токсично-сексично, токсично для личности, токсично для группы и угроза — выборка с подобным углубленным подходом к разметке токсичности на данный момент в российском сегменте интернета не представлено. Точность классификации составила от 75.2% до 97.7% по метке F1.

Также были спроектированы и разработаны Back-end и Front-end модули системы на языке программирования TypeScript, обеспечивающие функционал исследователя и эксперта по разметке .

В будущем будет проведен эксперимент по разметке актуального набора данных по сообщениям из социальной сети с использованием разработанного приложения, и сравнение полученных результатов с конкурентными приложениями.

СПИСОК ЛИТЕРАТУРЫ

1. Сергеева Ю. Вся статистика интернета и соцсетей на 2021 год — цифры и тренды в мире и в России / Ю. Сергеева. — Текст : электронный // WebCanape : официальный сайт. 2021. — URL: <https://www.web-canape.ru/business/vsya-statistika-interneta-i-socsetej-na-2021-god-cifry-i-trendy-v-mire-i-v-rossii/> (дата обращения: 01.04.2022).
2. Dolgushin M. Toxic Comment Classification Service in Social Network / M. Dolgushin, D. Ismakova, Y. Bidulya [et al.]. — DOI 10.1007/978-3-030-87802-3_15. — Text : electronic // Speech and Computer. — 2021. — Vol. 12997. — P. 157-165. — URL: https://link.springer.com/chapter/10.1007/978-3-030-87802-3_15 (date of application: 26.05.2022).
3. Toxic Comment Classification Challenge Overview / kaggle. — Text : electronic // kaggle.com = kaggle : [site]. — URL: <https://www.kaggle.com/competitions/jigsaw-toxic-comment-classification-challenge/overview> (date of application: 04.04.2022).

4. Куцев Р. А. Как организовать разметку данных для машинного обучения: методики и инструменты / Р. А. Куцев. — Текст : электронный // habr.com = Хабр : [сайт]. — URL: <https://habr.com/ru/post/572508/> (дата обращения: 04.04.2022).
5. CatBoost. Документация : [сайт]. — URL: <https://catboost.ai/> (дата обращения: 11.05.2022). — Текст : электронный.

Ю. А. АРКАДЬЕВА, И. Г. ЗАХАРОВА

Тюменский государственный университет, г. Тюмень

УДК 51(082)

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ПЛАНИРОВАНИЯ ВРЕМЕНИ РЕМОНТА ДВИГАТЕЛЕЙ НА МАШИНОСТРОИТЕЛЬНОМ ПРОИЗВОДСТВЕ

***Аннотация.** В статье рассмотрена проблема выбора информационного обеспечения для создания системы поддержки принятия решений при планировании времени, требующегося для ремонта двигателей на машиностроительном предприятии. Произведен анализ существующих решений и предложен новый подход к построению прогнозных моделей.*

***Ключевые слова:** производственное планирование, системы поддержки принятия решений, прогнозные модели, регрессионный анализ.*

Введение. Со времен зарождения промышленности до наших дней приоритетом в развитии любого производства являлось наращивание его темпов и увеличение объемов. Планирование — очень важный аспект в данном вопросе, ошибки в нем дорого обходятся предприятиям. Особенно отчетливо это заметно в отраслях высокого стратегического и экономического значения. Одной из таких является машиностроение и, в частности, двигателестроение.

Российская Федерация не так давно утвердилась в роли одного из лидеров мировой экономики. Поддержание подобного статуса обязывает участвовать в жесткой конкурентной борьбе с иностранными производителями в сфере создания двигателей [1].

Планирование вне зависимости от его вида и контекста применения основывается на некоторой оценке будущего состояния рассматриваемого объекта — прогноза. Точность этого прогноза