ТГ: Исмакова Даяна

ЦТ ДЛЯ ФИЛЬТРАЦИИ ТОКСИЧНЫХ   
КОММЕНТАРИЕВ В СОЦИАЛЬНЫХ СЕТЯХ

Тезаурус

А – администратор

К – комментарий

П – пользователь

ВВЕДЕНИЕ

Ручная фильтрация негативных комментариев (К) занимает много времени и вредит здоровью [1, 2], а автоматическая часто работает некорректно. Факторы замедляющие обработку К:

* использование замещающих символов, которые не способен определить компьютер;
* слова, которые вне контекста являются корректными;
* оскорбления, затрагивающие определенную группу людей без использования некультурной лексики.

**Проблема**: распространение токсичной лексики в социальных сетях и отсутствие ее быстрой и качественной модерации.

**Цель**: создать ЦТ,  повышающую скорость и качество обработки токсичных К.

1. **Сценарий «как есть»**

* Исполнитель: администратор (А).
* Предусловие: А авторизован, открыта социальная сеть.

Таблица . Основной процесс

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Общее название операции | Сценарии | Действие исполнителя | Отклик соцсети | Возможная программная реализация |
| 1 | Открытие окна | 1.1)Открытие К в группе | 1.1.а)Клик «открыть группу»  1.1.б) Клик «открыть пост»  1.1.в)Открыть секцию с К | Отображает комментарии | Открытие окна с К |
| 1.2)Открытие окна жалоб на конкретные К | 1.2.а)Клик «открыть окно жалоб» | Отображает окно с К |  |
| 2 | Фильтрация | 2.1) Распознавание НК | 2.1.а)Прочитывает К  2.1.б) Находит негативный К  2.1.в) Определяет степень токсичности |  | Предобработка текста[3], распознавание с помощью Байесовского классификатора |
|  |  | 2.2) Низкая степень токсичности К | 2.2.а) {2.1.а – 2.1.в}  2.2.б)Определил К как «Низкая степень токсичности»  2.2.в)Смотрит, есть ли данный П в списке нарушителей  2.2.в.1)Есть, но не превысил макс. количество раз, отправляет предупреждение П  2.2.в.2)Есть, превысил норму, блокировка П на неделю | Сохраняет П в список нарушителей |  |
|  |  | 2.3) Средняя степень токсичности К | 2.3.а) {2.1.а – 2.1.в}  2.3.б) Определил К как «Средняя степень токсичности»  2.3.в)Смотрит, есть ли данный П в списке нарушителей  2.3.в.1)Есть, но не превысил макс. количество раз, отправляет предупреждение П  2.3.в.2)Есть, превысил норму, П блокировка на 2 недели  2.3.г) Удаляет К | Сохраняет П в список нарушителей,  удаляет негативный К |  |
|  |  | 2.4) Высокая степень токсичности | 2.4.а) {2.1.а – 2.1.в}  2.4.б) Определил К как «Высокая степень токсичности»  2.4.в)Смотрит, был ли данный П в списке нарушителей  2.4.в.1) Был уж много раз, пожизненная блокировка  2.4.в.2) Был, блокировка на месяц  2.4.в.3) Не был, блокировка на 2 недели  2.5) Удаляет к | Сохраняет П в список нарушителей,  удаляет негативный К |  |
| 3 | Составление отчета | 3.1)За день | 3.1.а) Открывает окно статистики  3.1.б) Осматривает |  |  |
|  |  | 3.2)За неделю |  |  |  |
|  |  | 3.3)За месяц |  |  |  |
|  |  | 3.4)За все время |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

1. **Требования “клиента” к ЦТ**

Главная задача клиента разрешать конфликтные ситуации, т.е., когда система не способна определить является ли комментарий токсичным или нет. Желательно сократить эти конфликтные ситуации до минимума. 

1. Удобный и интуитивно понятный интерфейс
3. **Требования к ЦТ**

Разрыв страницы

ИСТОЧНИКИ

1. Victoria Flint, Темная сторона Google: проблемы модерации, URL: <https://dw.expert/2020/03/30/temnaya-storona-google-problemy-moderacii-eksklyuzivnoe-intervyu-s-dejzi-soderberg-rivkin/> (Дата обращения: 04.11.2020)
2. Николай Чумаков, Как работают модераторы «Вконтакте»: полная анонимность, стресс и тысячи постов в сутки, URL: <https://tjournal.ru/internet/88068-kak-rabotayut-moderatory-vkontakte-polnaya-anonimnost-stress-i-tysyachi-postov-v-sutki> (Дата обращения: 04.11.2020)
3. Wikipedia, Предварительная обработка данных, URL: <https://ru.wikipedia.org/wiki/Предварительная_обработка_данных> (Дата обращения: 10.11.2020)