

Детоксификация русскоязычных текстов

Ольга Пичужкина, Татьяна Гнедина

Задача

Написать модель для задачи автоматической детоксификации русскоязычных текстов.

Актуальность/мотивация:

Команда и роли

Ольга Пичужкина

- Сбор и препроцессинг данных

Татьяна Гнедина



Данные

<https://www.kaggle.com/blackmoon/russian-language-toxic-comments>

<https://www.kaggle.com/alexandersemiletov/toxic-russian-comments>

Всего

Baselines

- **Duplicate:** “наивный бейзлайн” — просто продублировать предложение без изменений;
- **Delete:** removal of rude and toxic from pre-defined **vocab**;
- **Retrieve:** retrieval based on cosine similarity between word embeddings from non-toxic part of **RuToxic** dataset;

Метрики оценки

План действий

- Собрать датасет комментариев на русском языке, размеченных как токсичные/нетоксичные
- Из полученного датасета

Список использованной литературы

- [Jin et al. Deep Learning for Text Style Transfer: A Survey](#) —
- [Dementieva et al. Methods for Detoxification of Texts for the Russian Language. Диалог-2021](#) (<https://github.com/skoltech-nlp/rudetoxifier>) —
отсюда взяты бейзлайны и метрики оценки.