

Техническая сторона
проекта

Задачи

- Выбор инструментов для краулинга: VK API
- Сбор материала при помощи парсинга;
- Проведение предобработки;
- Тематическое моделирование;
- Создание веб-интерфейса.

Методология

Предобработка:

- *Лемматизация с помощью `ru morphology2`*
- *Формирование словаря - минимальная частота 100*

*Алгоритм LDA из библиотеки `gensim`
(Coherence Score = 0.379)*

- *`update_every=1,`*
- *`chunksize=128,`*
- *`passes=50,`*
- *`alpha='auto',`*
- *`per_word_topics=True`*

LDA Mallet (Coherence Score 0.462)

REACTIONS PREDICTOR

POST TEXT

Predict

EXPECTED **No** VIEWS

EXPECTED **No** REACTIONS

Веб-интерфейс

- Загрузка модели тематического моделирования (LDA Mallet);
- Построение алгоритма предварительной обработки неизвестного текста;
- Построение алгоритма предсказания усреднённых значений по количеству просмотров и реакций в зависимости от темы.



<https://github.com/zozunin/bermuda>

Спасибо, что оставались с
нами :)