# Задание 4. Кластеризация

1. Сформулируйте предположение о том, на какие кластеры могут быть разделены ваши данные.
2. Постройте векторизатор ваших текстов с использованием характеристики TF-IDF.
3. Выведите топ-50 слов с наибольшим весом и с наименьшим. Как вы предполагаете, почему именно такие слова имеют наибольший и наименьший вес?
4. Выполните обучение кластеризатора:
   1. с использованием векторизация из прошлых заданий;
   2. с использованием характеристики TF-IDF;
   3. ДО предобработки и ПОСЛЕ предобработки.

## Ожидаемый результат

1. Файл jupyter-notebook, который содержит решение все перечисленных выше задач с поэтапным выводом результатов (как было в примере в лекции).
2. Текстовый документ с интерпретацией результатов - словесное описание результатов (дало ли результат изменение параметров).

## Дополнительно

Для лемматизации русских слов можно использовать библиотеку pymorphy2.