

Тема: СОЗДАНИЕ МЕТОДИКИ СЕМАНТИЧЕСКОГО
АНАЛИЗА ОТЗЫВОВ STEAM

Автор: Машина Юлия Дмитриевна, 7381

Руководитель: Борисенко Константин Алексеевич,
кандидат технических наук

Актуальность

Тематическое моделирование как метод семантического анализа помогает решать проблему ручного репетитивного анализа большого количества данных

Применение тематического моделирования в Quality Assurance видеоигровой индустрии поможет разработчикам в обработке обратной связи и поддержке пост-релизного продукта



Цель и задачи

Цель: анализ LDA, BERT, TF-IDF с помощью заранее выбранных метрик и произвести вывод полезной для аналитика информации.

Задача 1: Построить модель идентификации тем

Задача 2: Проанализировать результаты работы выбранных методов тематического моделирования

Задача 3: Произвести вывод полезной для аналитика информации

Задача 1

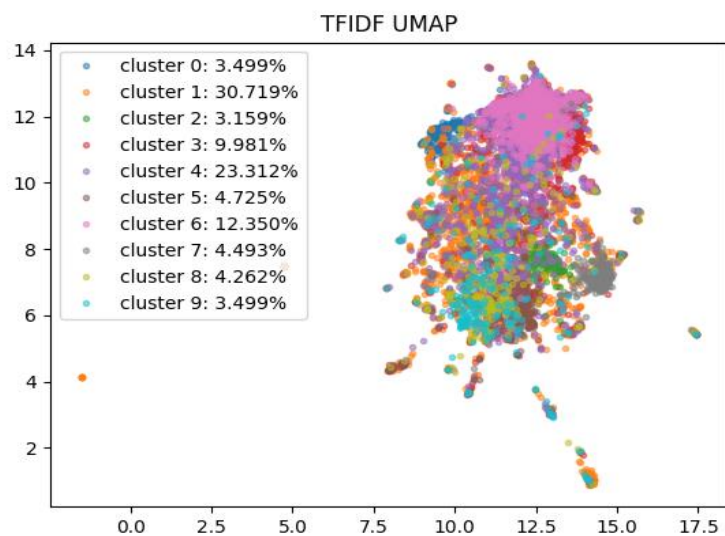
LDA - один из самых популярных методов тематического моделирования. Но в рамках достижения цели семантического анализа отзывов Steam присутствуют следующие минусы:

- Ненадежный результат при обработке коротких текстов
- Один отзыв \neq одна тема
- Базируется на bag-of-words, тем самым результат метода не зависит от контекста, тогда как значение отзывов от контекста зависит

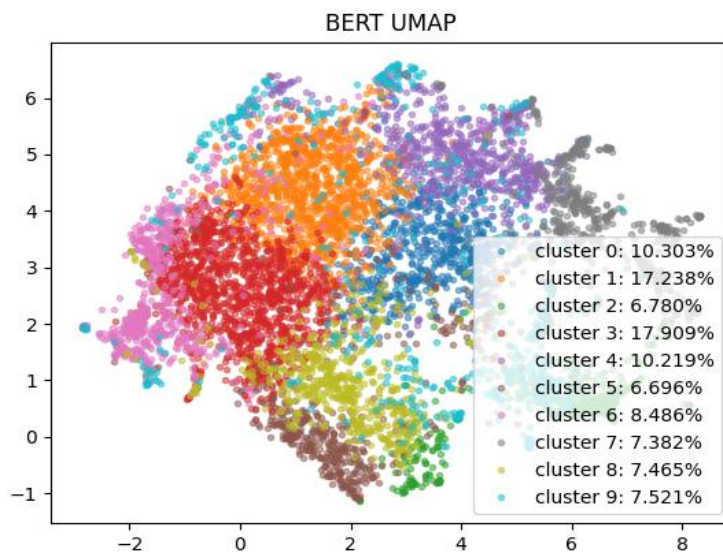
Помня об этих ограничениях, нам нужен метод, который включает полное содержание предложения, которое затем можно сгруппировать по аналогичным темам. Векторизация вместе с кластеризацией представляется самым прямолинейным решением.

Задача 1

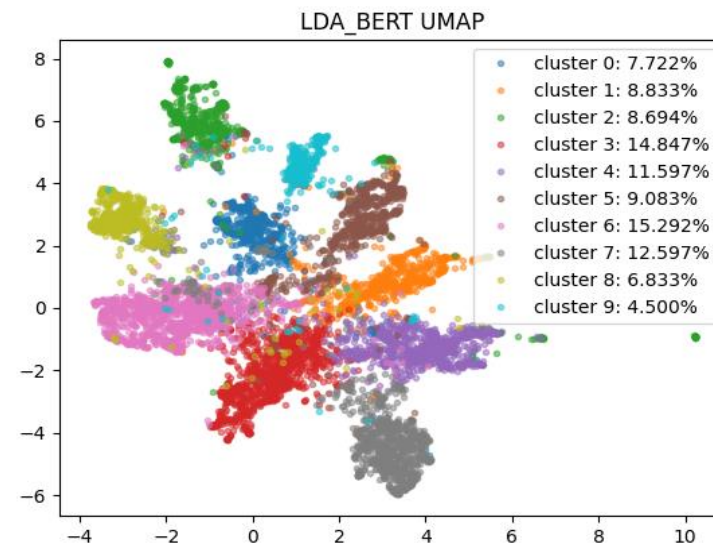
TF-IDF



BERT



LDA+BERT



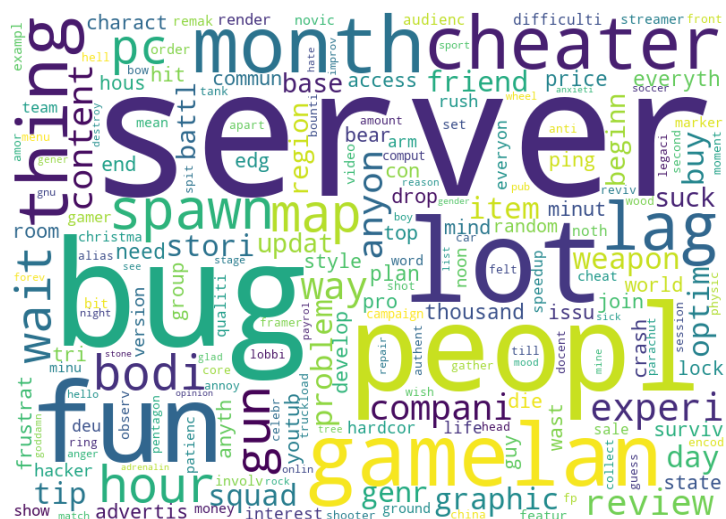
Задача 2

Лучшие показатели принадлежат LDA+BERT

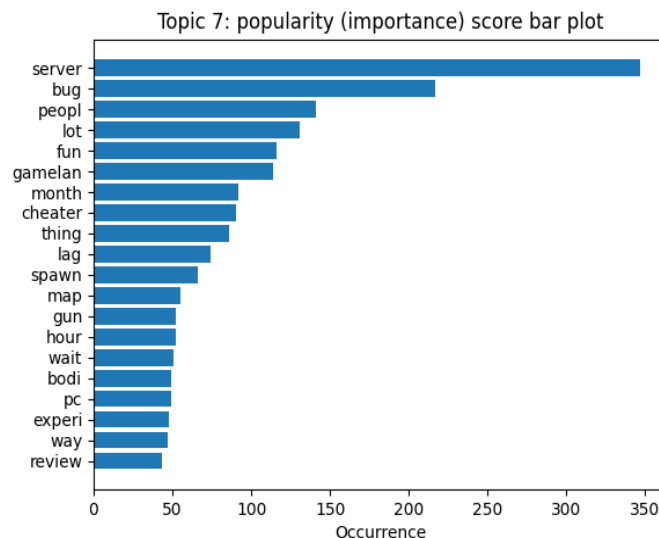
Метрика/ метод	LDA	BERT с кластеризацией	TF-IDF с кластеризацией	LDA+BERT с кластеризацией
Мера связности (метод u_mass)	-5.234	-4.372	-2.161	-3.399
Мера связности (метод c_v)	0.483	0.543	0.538	0.551
Оценка силуэта	—	0.064	0.025	0.234

Задача 3

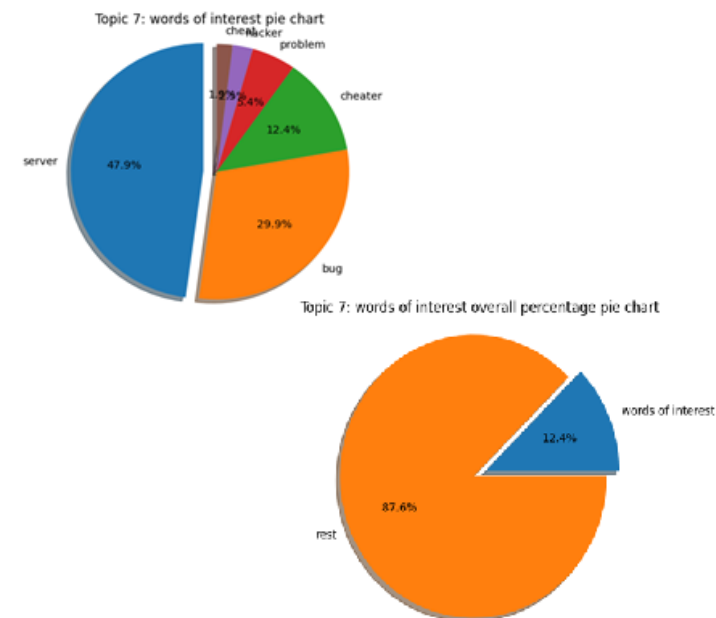
Облако слов



Столбчатая диаграмма



Круговые диаграммы



Задача 3

Примеры отзывов

"first game that i enjoyed from that genre so i guess thats a plus v really good and fun there are some bugs though but probably going to be fixed Huge lags for the first minute after parachuting down after that it fixes Sometimes just bugs out and cant revive or heal yourself but happened once couldve been servers problemall in all its an early access just report bugs to developer and cry if its not fixed next month not instantly"

- Пример отзыва для темы проблем с сервером и интересующими словами bug, lag

Апробация работы

Ссылка на работающее приложение: coming soon

Заключение

Задача 1: построена модель тематического моделирования

Задача 2: в ходе анализа методов был выделен наиболее подходящий согласно своим показателям – LDA+BERT

Задача 3: выводимая информация включает в себя облако слов, столбчатую и круговые диаграммы, а также примеры отзывов на интересующую тему с интересующими словами.

Была сформирована методика

