

Частотный анализ лексики новостных Telegram-каналов Владивостока и Сахалина

Подсвирова Полина, Трякина Екатерина

Цель проекта

Выявить лексические и тематические особенности постов новостных Telegram-каналов Владивостока и Сахалина и Курил на основе частотного анализа лемматизированных текстов.

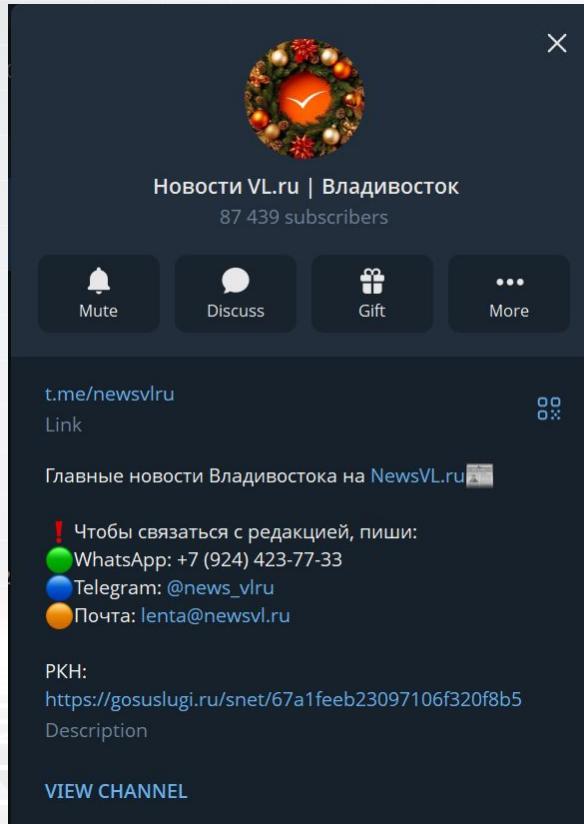
Задачи проекта

1. Выбрать тг-каналы
2. Отобрать тексты
3. Нормализовать, токенизировать и лемматизировать тексты
4. Создать униграммные, биграммные и триграммные моделей
5. Выявить лексические и тематические особенности
6. Провести сопоставительный анализ

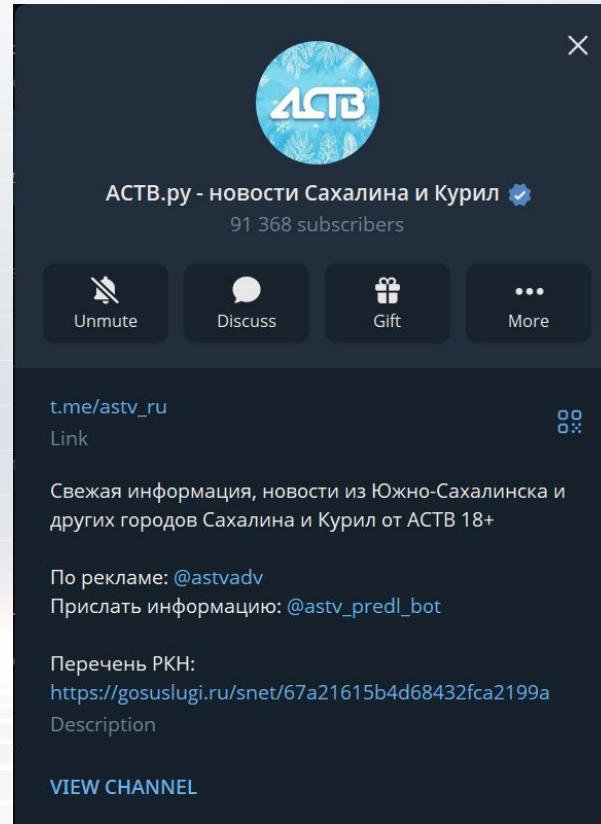
1 этап

Выбор тг-каналов

Новостной канал Владивостока vl.ru



Новостной канал Сахалина и Курил astv.ru

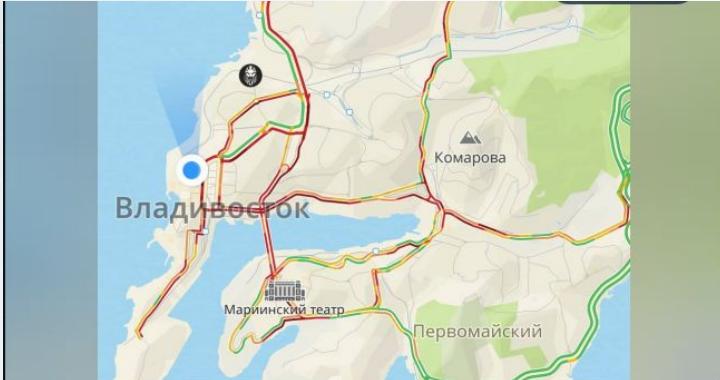


VS

2 этап

Отбор постов

Период: 31 декабря 2024 год - 7 января 2026 год



10 из 10 – вечерние пробки снова парализовали Владивосток

Жители Владивостока начинают и заканчивают уже второй рабочий день 2026 года в пробках. Городские службы всё ещё пытаются убрать снег, но получается это не везде. А горожане не могут сменить личные автомобили на общественный транспорт, которого не дождёшься – поэтому едут на работу и притыкают машины в наиболее удобный сугроб. В итоге дороги становятся настолько узкими, что не везде могут разъехаться даже две легковушки.



Морозы усилиятся: погода на Сахалине и Курилах на 6 января

Вторник на островах будет малоснежным. В большинстве районов ясно, однако столбик термометра продолжит падать вниз - кое-где он достигнет отметки в -24 °С.

Astv.ru составил подробный прогноз в районах области на основе данных openweathermap.org 

ACTB.py - новости Сахалина и Курил

91 538 subscribers

Pinned message

Информатор astv.ru теперь и в MAX: куда и как отправить обращение в редакцию

⚡⚡ Цены на топливо взлетели в Южно-Сахалинске

Стоимость поднялась в ночь на 2 января, сообщают автомобилисты. Ценник на все виды горючего подорожал почти на рубль и теперь составляет:

- АИ-92 — 69.78₽
- АИ-95 — 71.99₽
- АИ-98 — 99.63₽
- ДТ — 90.67₽

В последний раз ННК «радовала» сахалинцев повышением в канун Нового года, 31 декабря.

Прислать новость

Подписывайся на нас:
MAX | Telegram | WhatsApp | ВК

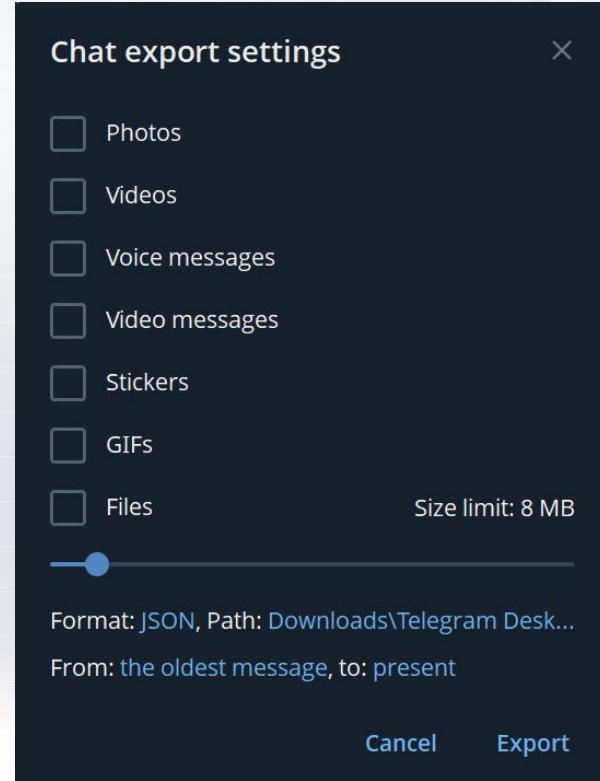
515 39 11 11 5

24.6K edited 11:29

171 comments >

101

ACTB.py - новости Сахалина и Курил



```
1 {
2   "name": "Новости VL.ru | Владивосток",
3   "type": "public_channel",
4   "id": 1040881583,
5   "messages": [
6     {
7       "id": 99795,
8       "type": "message",
9       "date": "2024-12-31T02:38:54",
10      "date_unixtime": "1735576734",
11      "edited": "2024-12-31T02:40:09",
12      "edited_unixtime": "1735576809",
13      "from": "Новости VL.ru | Владивосток",
14      "from_id": "channel1040881583",
15      "reply_to_message_id": 99791,
16      "file": "(File not included. Change data exporting settings to download.)",
17      "file_size": 27357587,
18      "thumbnail": "(File not included. Change data exporting settings to download.)",
19      "thumbnail_file_size": 17623,
20      "media_type": "video_file",
21      "mime_type": "video/mp4",
22      "duration_seconds": 76,
23      "width": 1280,
24      "height": 720,
25      "text": "Во Владивостоке припорошило и любители подрифтить тут же захватили дороги города.",
26      "text_entities": [
27        {
28          "type": "plain",
29          "text": "Во Владивостоке припорошило и любители подрифтить тут же захватили дороги города."
30        }
31      ],
32      "reactions": [
33        {
34          "type": "emoji",
35          "count": 231,
36          "emoji": "\ud83d\udcbb"
```

3 этап

Обработка текста

Обработка текста

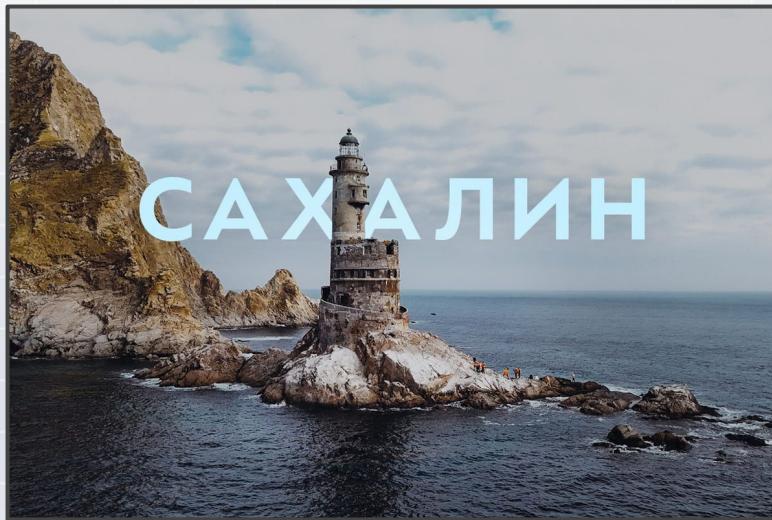
Очищение от ссылок,
эмодзи, служебных
символов, знаков
препинания

Удаление стоп-слов +
“Подпишись” и тд





ВЛАДИВОСТОК



САХАЛИН

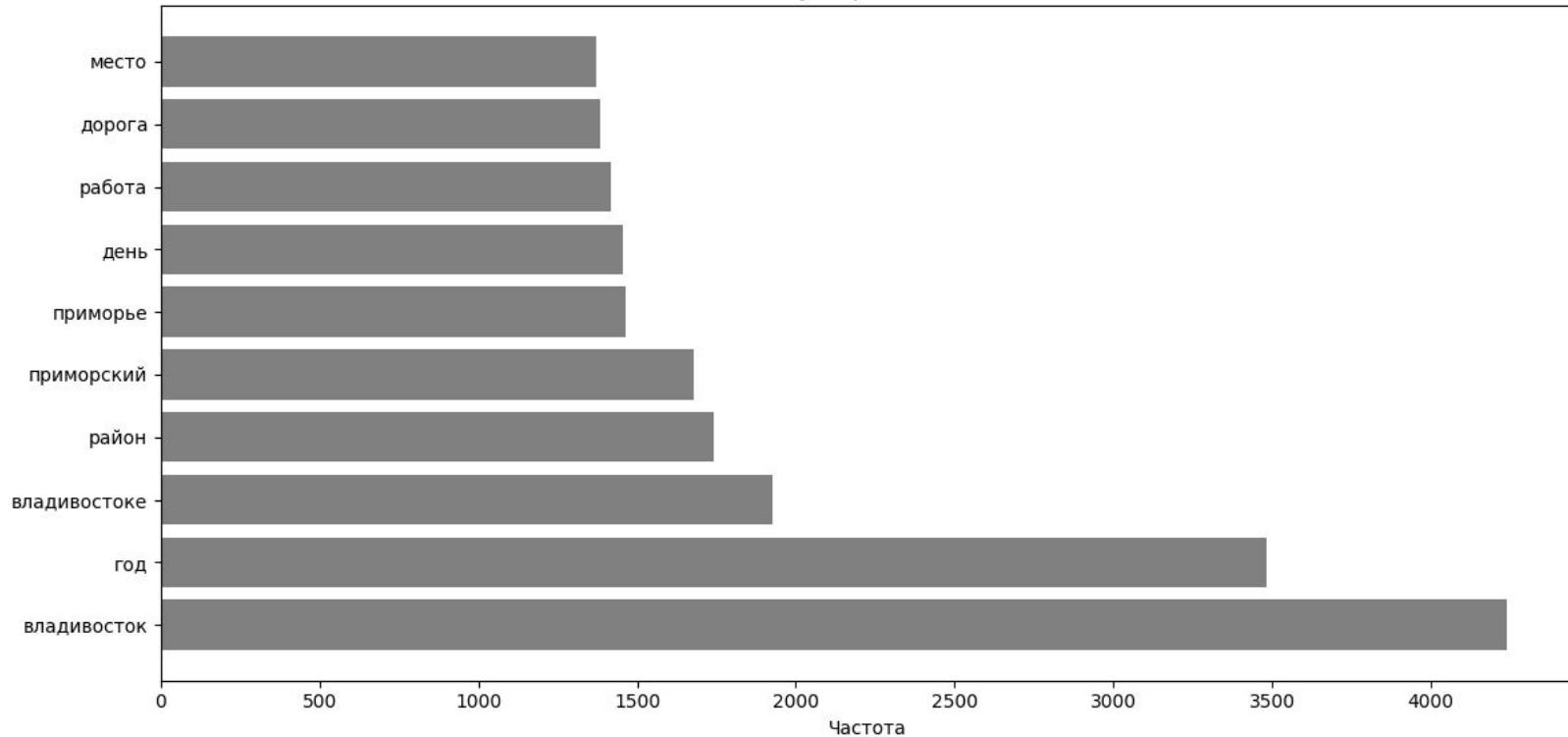
9936
текстов

10 578
текстов

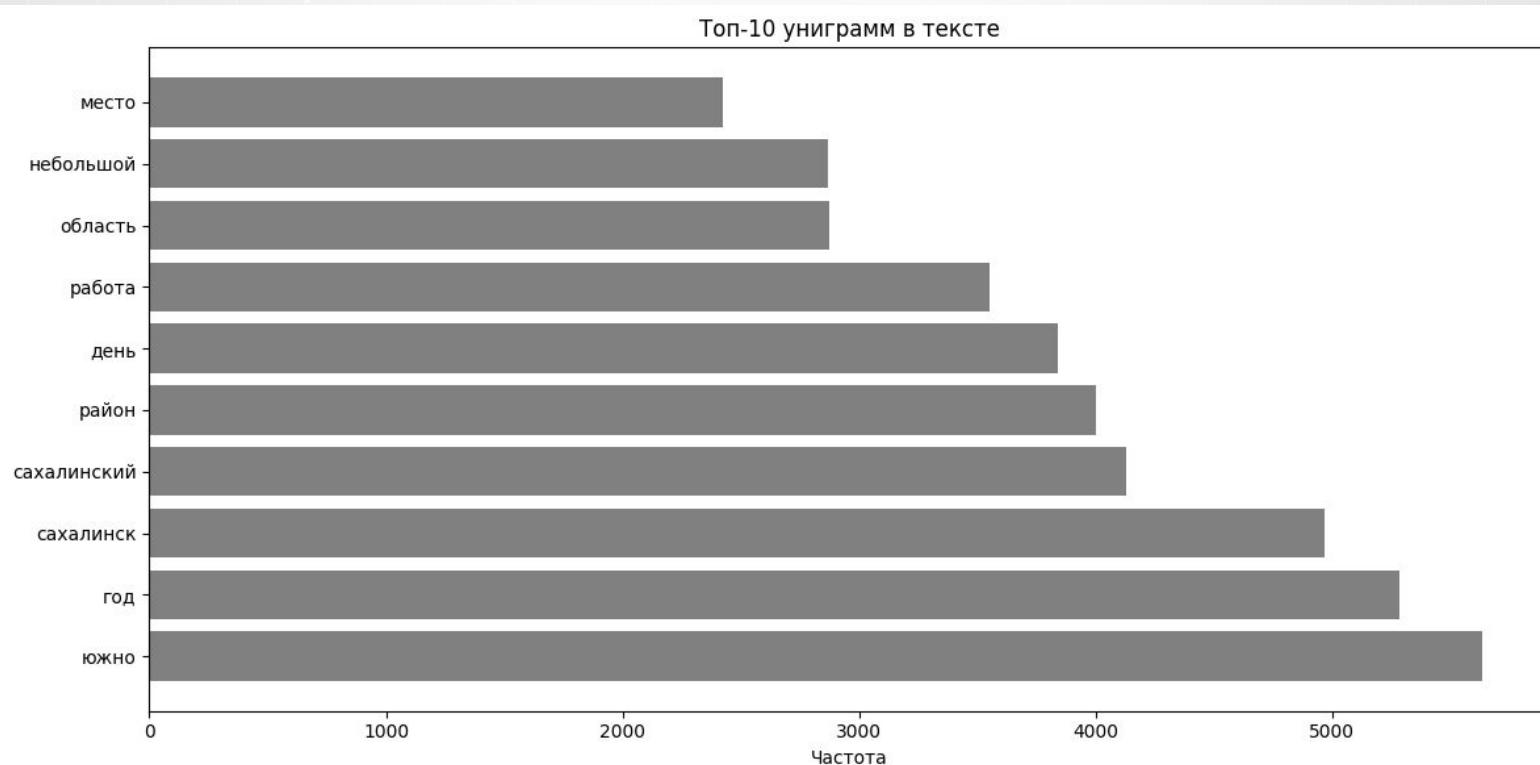
4 этап

Создание n-грамм

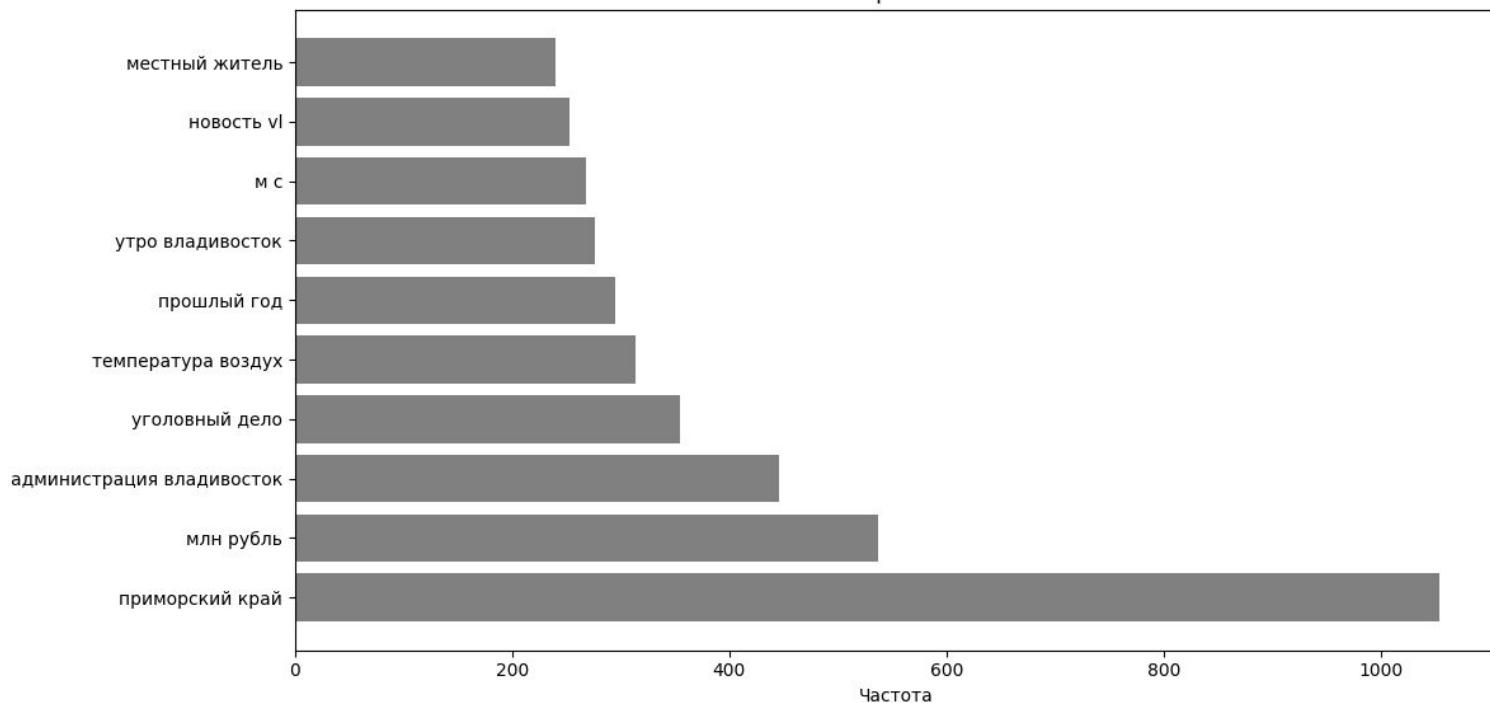
Топ-10 униграмм в тексте



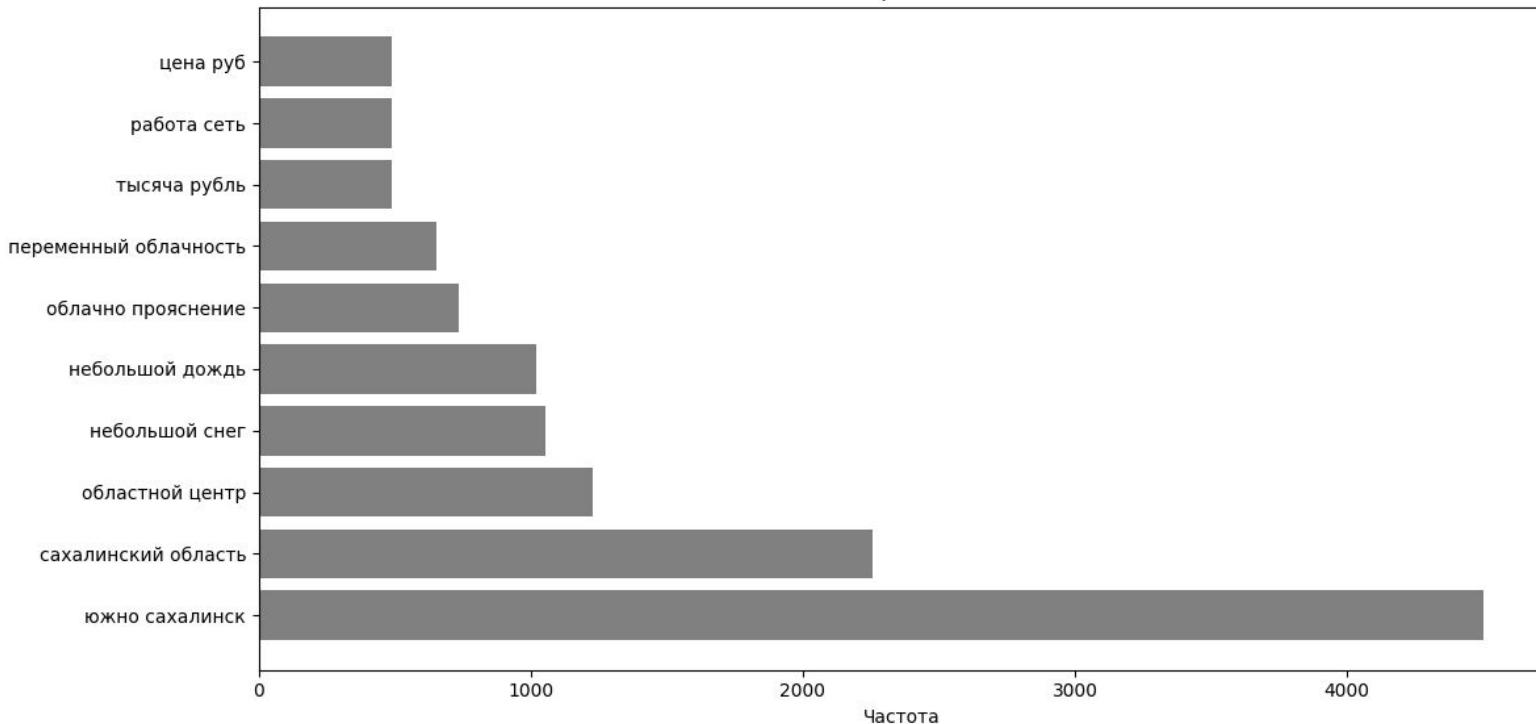
Топ-10 униграмм в тексте



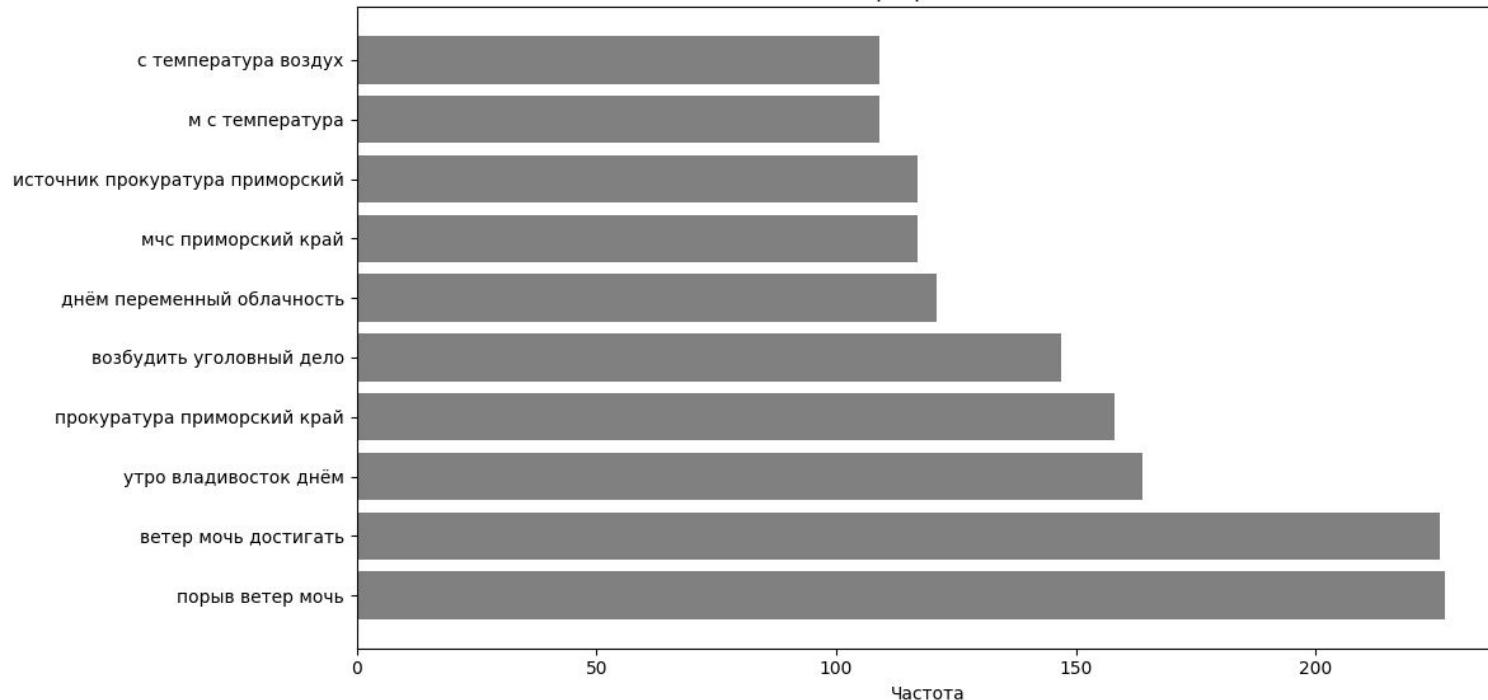
Топ-10 биграмм в тексте



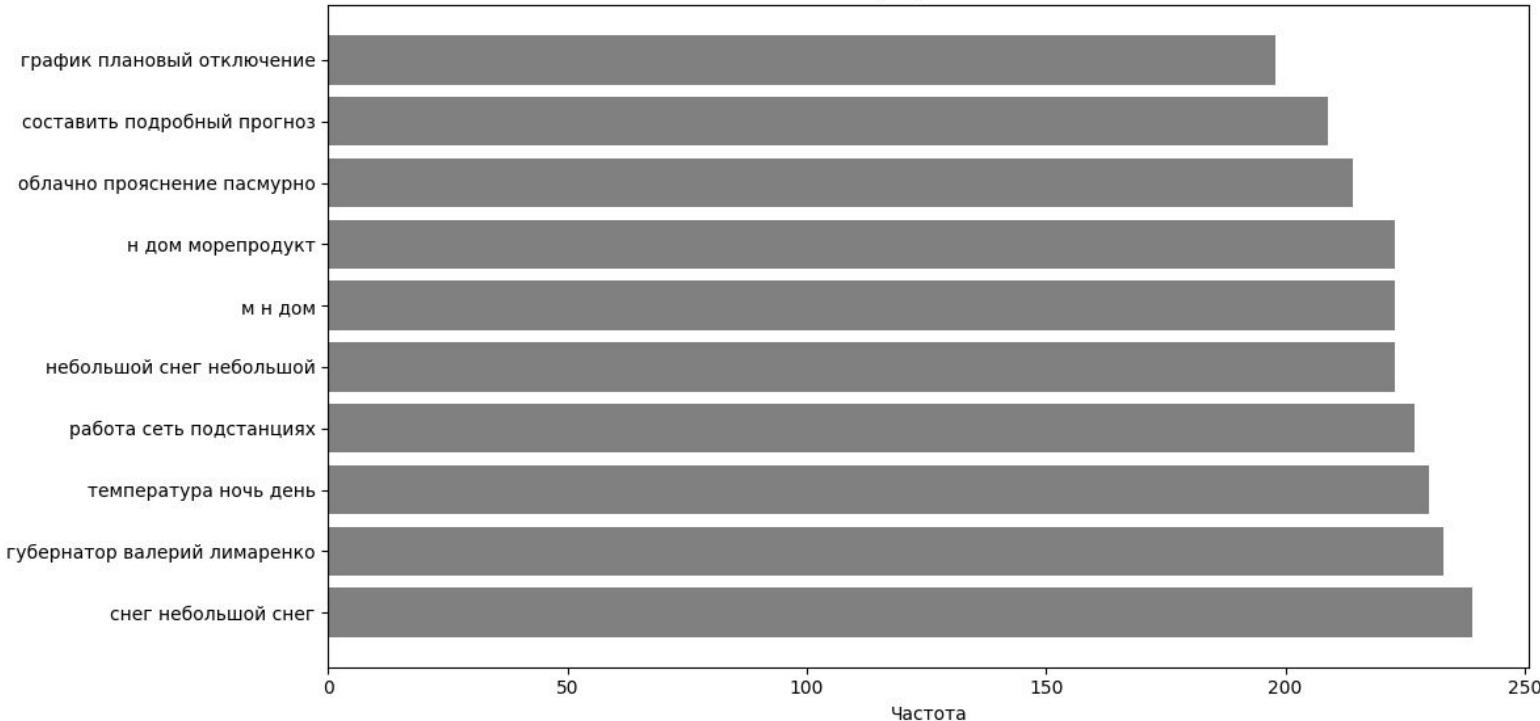
Топ-10 биграмм в тексте



Топ-10 триграмм в тексте



Топ-10 триграмм в тексте



5 этап

Тематические особенности

Погода



Воздух прогреться
Облачность осадок
Переменный облачность
Порыв ветер
14 м с
Ветер мочь достигать



Дороги

Автомобиль
Машина
Дорога
Транспортные средство
Проезжий часть

Места



Снеговой падь
Океанский проспект
Русский остров
Набережная спортивный
гавань



Уникальные

Работа
Квадратный метр
Возбудить уголовный дело
МЧС приморский край
Восточный экономический форум



Погода

Небольшой облачность
Облачный прояснение
Небольшой дождь
Небольшой снег



ЖКХ

Плановый отключение
РВК сахалин
Островной энергетики снова
Обесточить десяток жилой



Места

Город
Село
Дом морепродукт
Десктай сад
Санаторий
Областной центр



Уникальные

Южно-сахалинск
Островной столица
Южно-курильск
Губернатор Валерий
Лимаренко

6 этап

Сопоставительный анализ

Сопоставительный анализ

1

Общее семантическое поле “погода”, но в нем разные лексические единицы

2

Региональные проблемы:
Дороги vs ЖКХ

3

Характерные для региона особенности

Выводы

1. Нет универсальной функции для очистки telegram-постов
2. Новостные каналы затрагивают примерно одни и те же темы (проблемы, происшествия, текущая ситуация в регионе)
3. Биграммы и триграммы позволяют отслеживать не просто отдельные понятия, а дискурсивно-специфичную лексику