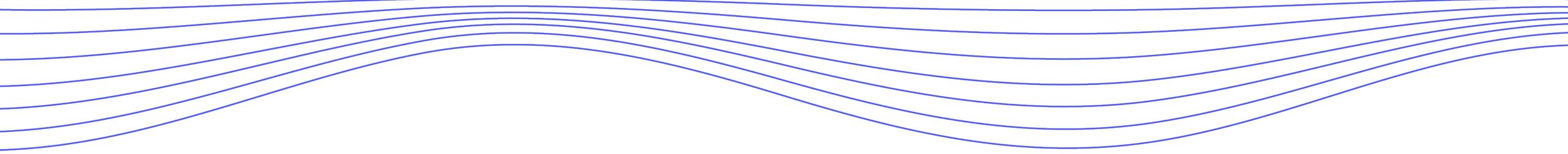




Social issue is just a

World  
AI & Data  
Challenge

МЕЖДУНАРОДНЫЙ КОНКУРС  
ПО РЕШЕНИЮ ГЛОБАЛЬНЫХ  
СОЦИАЛЬНЫХ ЗАДАЧ И ВНЕДРЕНИЮ ИХ  
РЕШЕНИЙ ПРИ ПОДДЕРЖКЕ  
МЕЖДУНАРОДНЫХ ЭКСПЕРТОВ

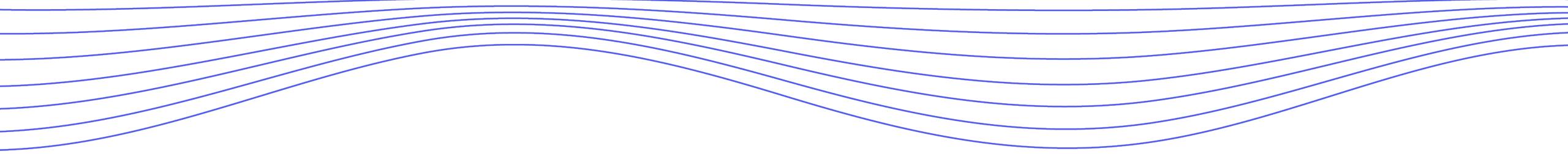


# **FreyrMonitoring**

**Шаталин Игорь**

<https://git.asi.ru/shatalin.ip/freyrmonitoring>  
<http://84.201.164.182:5000>

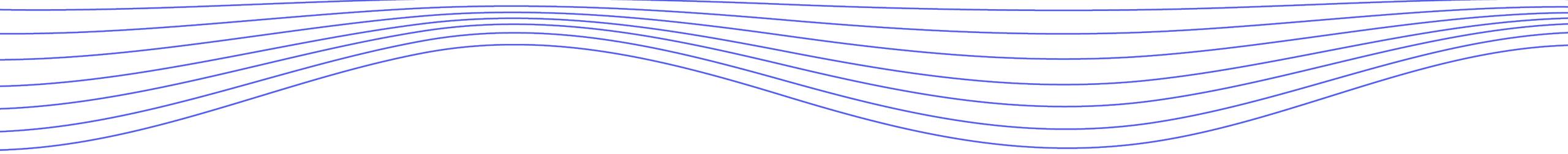
Индекс одобрения действий органов власти  
Нижний Новгород



Цель — выявить не просто все сообщения, связанные с регионом, а те, которые отражают отношение к региональной власти

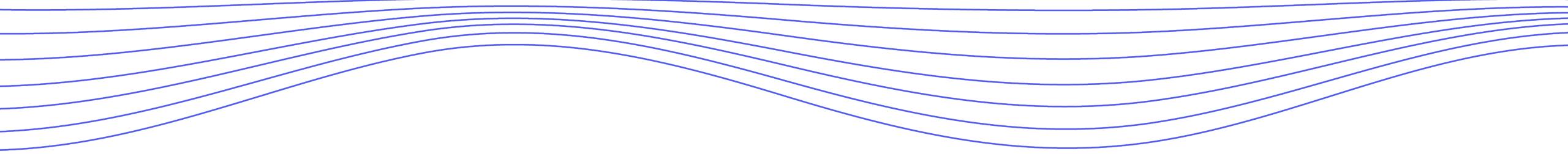
Поэтому наша команда решила организовать сбор и систематизацию данных следующим образом:

- **Парсинг региональных ресурсов.**
- **Отбор** из всех материалов только тех, которые подходят нам: выражающие отношение к власти, освещдающие злободневные темы региона.
- **Определение тональности высказываний** (автор положительно освящает событие, отрицательно или нейтрально).



Чтобы заказчику удобно было пользоваться собранными материалами и делать определённые выводы, все собранные материалы группируются по 3 категориям:

- **Группы сообщений по персоналиям.** Отношение жителей к депутату или первому лицу региона как правило отражает их отношение к деятельности региональной власти в целом. Поэтому система группирует сообщения по наиболее популярным персонам.
- **Группы сообщений по организациям.** Иногда жители бывают возмущены бездействием какого-то ведомства или деятельностью отдельной организации. Поэтому система собирает группы сообщений по наиболее обсуждаемым в сети организациям.
- **Группы сообщений по темам.** Сообщения кластеризуются на наиболее популярные темы. Это поможет заказчику понять, какие вопросы на данный момент являются актуальными для жителей региона.



## Используемые технологии

Помимо краулинга, сервис активно использует в своей работе методы машинного обучения, поэтому наш выбор был стандартен: Ubuntu + Python + PostgreSQL + Django.

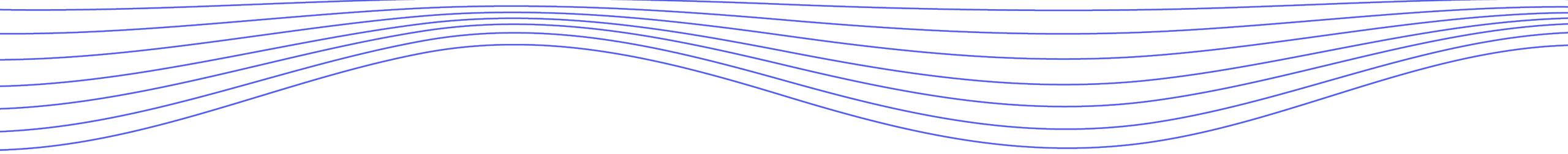
Кроме того, мы использовали такие популярные и хорошо зарекомендовавшие себя открытые продукты, как Celery, Redis, Pytorch, Transformers, Natasha, Kaldi.

Все использованные технологии распространяются под открытыми лицензиями.

# Ютуб и Телеграм

Как правило, сервисы анализа мнений в интернете индексируют лишь веб-страницы. Но общество не стоит на месте, люди пользуются и другими средствами распространения информации. Поэтому мы решили расширить наши парсеры, и создали, помимо краулеров веб-страниц сайтов и соцсетей, 2 тестовых:

- **Краулер телеграм-групп.** Телеграм стал очень популярным в России, поэтому обходить его просто нельзя.
- **Краулер видеороликов с Ютуба.** Ютуберы высказывают своё мнение в видеороликах, к которому прислушиваются многие, но в текстовом виде этой информации нет. Поэтому мы создали специальный краулер, который распознаёт речь в видеоролике для дальнейшего её анализа.



# Дальнейшее развитие

Сейчас наш сервис имеет минимальную версию для ознакомления. Он содержит набор краулеров лишь для небольшого количества сайтов Нижнего Новгорода. Однако, мы нацелены на то, чтобы в ходе внедрения довести сбор данных до совершенства, расширив набор инструментов для парсинга.

Но для тестирования наш код уже сейчас может использовать любой регион. Ряд краулеров обладает универсальностью. Это, например, краулеры для соцсетей: "Вконтакте", "Инстаграм", "Одноклассники". И краулеры для региональных версий общероссийских сайтов: "Коммерсант", "Вести", "Московский комсомолец" и др.

Методы же классификации и кластеризации собранных материалов универсальны и могут использоваться без доработок любым регионом.

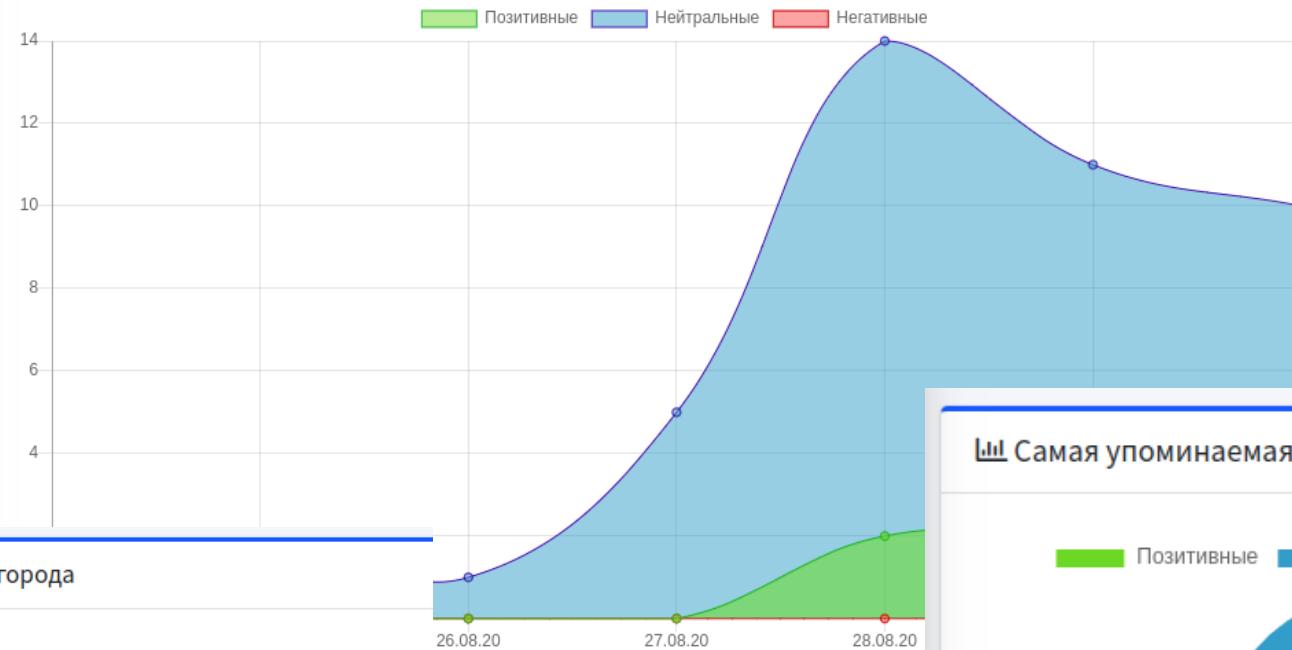
# Интерфейс

## Самая популярная тема: Бюджет Нижнего Новгорода

Всего **45** статей и постов на эту тему. И вот некоторые заголовки:

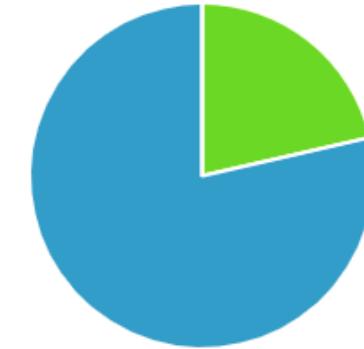
- Дефицит бюджета Нижнего Новгорода может увеличиться на четверть
- В Нижнем Новгороде состоялся мотосубботник
- Нижегородская мэрия предложила увеличить дефицит бюджета на 253 млн рублей
- 23 млн рублей потратят на средства защиты во время выборов в Нижнем Новгороде
- Более 330 млн рублей получит Нижний Новгород на строительство детсадов

## Статистика проиндексированных материалов за 7 дней

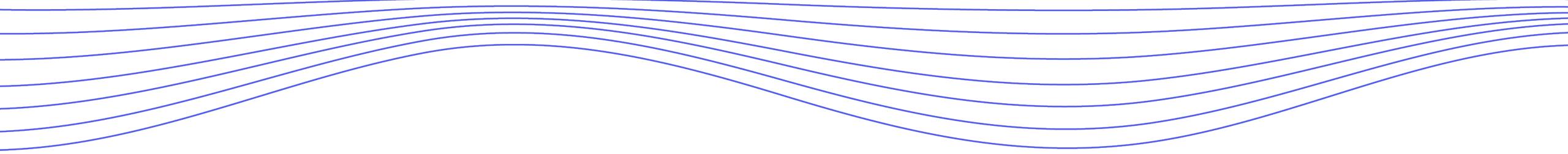


## Самая упоминаемая персона: ГЛЕБ НИКИТИН

Позитивные    Нейтральные    Негативные



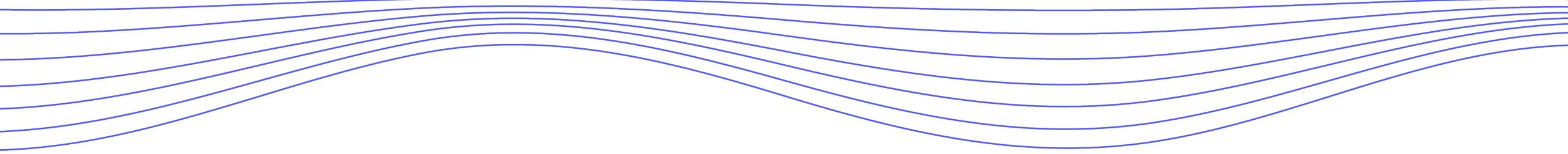
Демо-версия: <http://84.201.164.182:5000>



Наша команда планирует активнейшем образом участвовать в дальнейшем развитии проекта. И готова с радостью помочь любому региону во внедрении!

## Состав Команды

Шаталин Игорь  
[shatalin.ip@gmail.com](mailto:shatalin.ip@gmail.com)



**Спасибо за внимание!**

Шаталин Игорь

[shatalin.ip@gmail.com](mailto:shatalin.ip@gmail.com)

Демо-версия: <http://84.201.164.182:5000>

The background of the entire image is a dark, slightly blurred photograph of a city skyline filled with numerous skyscrapers, their windows forming a complex grid pattern.

D | I | T | I  
M A S T E R S

[datamasters.ru](http://datamasters.ru)