**Руководство пользователя Клиентского интерфейса мониторинга уровня социальной напряженности Система анализа открытых данных (далее – Система)**

**Оглавление**

[1. **Общие сведения** 2](#_Toc82697133)

[1.1. Общие сведения о Системе 2](#_Toc82697134)

[1.2. Источники информации 2](#_Toc82697135)

[1.3. Принцип проведения анализа 3](#_Toc82697136)

[2. **Как начать работу с Системой** 5](#_Toc82697137)

[2.1. Технические требования 5](#_Toc82697138)

[2.2. Вход в Систему 6](#_Toc82697139)

[3. **Интерфейс системы** 6](#_Toc82697140)

[3.1. Общий вид 6](#_Toc82697141)

[3.2. Перечень основных вкладок 7](#_Toc82697142)

[3.3. Перечень основных показателей на вкладках 10](#_Toc82697143)

[**4. Принципы сортировки результатов анализа 11**](#_Toc82697144)

[4.1. Фильтры по району 12](#_Toc82697145)

[4.2. Фильтры по сфере деятельности 13](#_Toc82697146)

[4.3. Фильтры по типу занятости 14](#_Toc82697147)

[4.4. Фильтры по уровню дохода 15](#_Toc82697148)

[4.5. Фильтры по возрасту 15](#_Toc82697149)

[4.6. Фильтры по источнику 15](#_Toc82697150)

[5. **Принципы распределения результатов анализа** 16](#_Toc82697151)

[5.1. Распределение данных по тематикам и факторам 16](#_Toc82697152)

[5.2. Распределение уровня отношения и сообщений по времени 16](#_Toc82697153)

[5.3. Распределение уровня отношения по районам региона 17](#_Toc82697154)

[5.4. Распределение уровня отношения по районам города 17](#_Toc82697155)

[5.5. Распределение количество людей, охвата и потенциального охвата сообщений по времени 18](#_Toc82697156)

[6. **Экспорт данных** 18](#_Toc82697157)

[6.1. Экспорт распределения по тематикам и факторам 18](#_Toc82697158)

[6.2. Экспорт графиков и диаграмм 19](#_Toc82697159)

# Общие сведения

## Общие сведения о Системе

Система представляет из себя механизм, который находится на стыке программирования и нейробиологии, который собирает, обрабатывает, классифицирует и хранит информацию из всех возможных открытых источников в автоматическом режиме.

В основе системы лежит технология – собственная исключительная разработка, которая дорабатывалась, обучалась и тестировалась в сфере государственного и муниципального управления в течение более чем 7 лет. Структурно состоит из 3 уникальных технологических модулей в виде обученных на большом массиве информации нейронных сетей, функционирующих во взаимосвязи друг с другом: модуль распознавания естественного языка, модуль лингво-морфического анализа и модуль профилирования. Технология имеет под собой большую научно-техническую базу, подтвержденную множеством патентов и научных публикаций. На сегодняшний день система была апробирована и применяется в 7 субъектах Российской Федерации.

## Источники информации

Система оперирует данными из всех возможных открытых источников данных. В соответствии с заданными критериями поиска, призванными ограничить территорию или целевую аудиторию (далее – ЦА) система в автоматическом режиме производит поиск и сбор информации из открытых источников, таких как:

* Все популярные социальные сети (Вконтакте, одноклассники, facebook и др.);
* Более 50 000 интернет СМИ и новостных агрегаторов (РБК, ТАСС, Известия, Ведомости, Яндекс.Новости, Yahoo News и др.);
* Публичные чаты и каналы в мессенджерах (Whatsapp, telegram, Viber);
* Торговые онлайн-площадки (Алиэкспресс, Wildberries, Avito, Яндекс.Маркет и др.);
* Площадки-отзовики (Яндекс, Google, ireccomended.ru, otzovik.com и др.);
* Форумы и интернет сообщества (Пикабу, 2ch, womans.ru, Яндекс.Дзен, Fishki.net);
* Сайты с ответами на вопросы (Ответы mail.ru, Яндекс.Q, quora и др.);
* Геоинформационные сервисы (Яндекс.Карты, Google Maps, 2Gis, Wikimapia).
* Видеохостинги (Youtube, Rutube и др)
* Любые другие открытые источники данных (по запросу в анализ и мониторинг может быть добавлен практически любой источник)

## Принцип проведения анализа

Для более полного понимания того, что позволяет и как работает Система, следует рассмотреть основной принцип проведения анализа. Рассмотрим этот принцип поэтапно:

* Идентификация в сети и сбор информации о требуемой ЦА.

В рамках проведения первичного анализа ЦА происходит анализ связанности всех определенных в сети аккаунтов с целью выявление уникальных аккаунтов и составления их лингво-морфических портретов.

Следующей стадией при выполнении анализа является определение тональности того контента, который генерируются уникальными аккаунтами из числа ЦА, а также выявление основных факторов, оказывающих наибольшее влияние на пользователей в сети.

Последующая аналитика выполняется с разбивкой на заранее установленные Заказчиком сферы или тематики. Также возможно выявление основных негативно и позитивно настроенных пользователей и лидеров мнений по каждой из тематик.

* **Полный анализ собранного массива информации, анализ поведения ЦА, выявление и оценка факторов, оказывающих наибольшее влияние на ЦА, а также выявление основных лидеров мнений и построение социальных графов и графов движения информации.**

В зависимости от общих объединяющих признаков по результатам сбора информации будет произведена сегментация ЦА на подгруппы. В рамках исследования поведения и предпочтений пользователей выстраиваются социальные графы – взаимосвязи и взаимодействия внутри ЦА. Построение социальных графов позволяет лучше понять ЦА и сегментировать ее на множество подгрупп для выполнения дальнейшей, более глубокой аналитики, а также для выполнения предиктивного анализа распространения информации.

Построение графов движения информации также позволяет визуализировать и детально проанализировать маршрут распространения информации: определить первоисточник, выстроить цепочку дальнейшего распространения информации, определить точки, в которых информация меняет свою тональность.

Понимание процесса хождения информации в сети позволяет выстроить алгоритмы наиболее эффективного взаимодействия не только со всей ЦА, но и с отдельными ее подгруппами.

При наложении графа социальных связей на граф движения информации в сети всю анализируемую ЦА можно разделить на взаимосвязанные друг с другом подгруппы, исследовать факторы, оказывающие влияние на каждую из групп, выделить лидеров мнений в каждой группе, а также отследить процесс обмена информацией между данными группами.

С практической точки зрения данный механизм позволяет определить те точки и рычаги, воздействуя на которые, можно управлять как мнением отдельных подгрупп, так и всей ЦА в целом. Это становится возможным именно благодаря знанию того, кто чем интересуется, какие у кого основные источники потребления информации и кто для кого является лидером мнений

* **Определение географической локализации пользователей, выстраивание маршрутов их перемещения в течение будних/выходных дней, их отношения к данному объекту.**

По результатам кроссмэтчинга аккаунтов и проведения анализа всей информации, написанной в сети уникальным аккаунтом, также становится возможным:

* локализовать пользователя до квадрата в   
  500 х 500 метров на карте;
* построить маршрут перемещения пользователей (в будние и выходные дни);
* **Визуализация полученных результатов многофакторного анализа ЦА.** Это достигается за счет настраивания интерактивного дашборда, на котором будут представлены ключевые результаты анализа, а также на базе которого будет осуществляться регулярный мониторинг ситуации с обновлением полученных результатов первичного анализа.

После проведения первичного исследования ЦА, возможно регулярное обновление полученных результатов благодаря постоянно работающему модулю парсинга, задачей которого является постоянный сбор вновь поступающей информации. Результаты мониторинга и анализа выводятся на интерактивном дэшборде. Отображаемые метрики на дэшборде настраиваются индивидуально в зависимости от запроса заказчика.

Более того, для оперативного информирования заказчика об изменениях или новых инцидентах настраивается система автоматического информирования, которая может существовать как в виде рассылок по электронной почте, телеграм-бота или же интегрироваться с внутренними системами заказчика по API.

* **Детектирование информационных атак и управление репутационными рисками**

Параллельно процессу анализа поведения пользователей в сети система в автоматическом режиме выполняет анализ всех аккаунтов по более чем 20 метрикам, часть из которых формируется посредством лингвистического анализа. В результате данного анализа происходит идентификация и кроссмэтчинг аккаунтов и определяются так называемые «бот аккаунты», которыми управляют иные люди, нежели это было заявлено в профиле и названии аккаунта.

Используя как собственные технологические решения, так и встроенные алгоритмы социальных сетей, становится возможным оперативно отрабатывать информационные атаки либо блокируя аккаунты-источники, либо автоматически понижая охват фейковых публикаций и комментариев.

Также в случае обнаружения критических/негативных всплесков, возможна разработка уникальных технологических решений по управлению возникшим негативным фоном или дискуссией.

# Как начать работу с Системой

## Технические требования

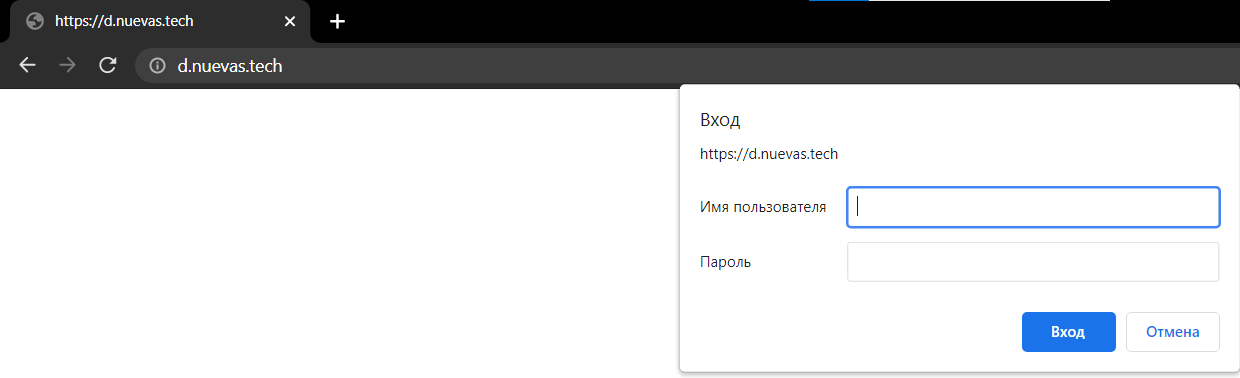
**Технические требования ПО:** операционная система семейства Unix; веб-сервер "Apache HTTPD version >2.2.15"

**Технические требования на стороне клиента:**

* Интернет соединение на скорости не менее 5 Мб/сек
* Операционная система семейства Microsoft Windows (Vista, 7, 8, 10) или семейства Mac OS X;
* Браузер Microsoft Internet Explorer 9.0 и выше или Mozilla Firefox, Google Chrome, Safari, Opera;
* Пакет Windows Media Player 7.1 версии и выше, содержащие кодеки для просмотра видеофрагментов Windows Media Codec 9, или проигрыватель медиафайлов VLC;
* Adobe Acrobat Reader для чтения PDF документов;
* MS Office Word и MS Office Excel, или другой продукт, поддерживающий форматы MS Office 97-2010.

## Вход в Систему

Для входа в систему нужно открыть окно браузера и перейти по ссылке <https://d.nuevas.tech/>. Далее необходимо ввести предоставленный логин и пароль:

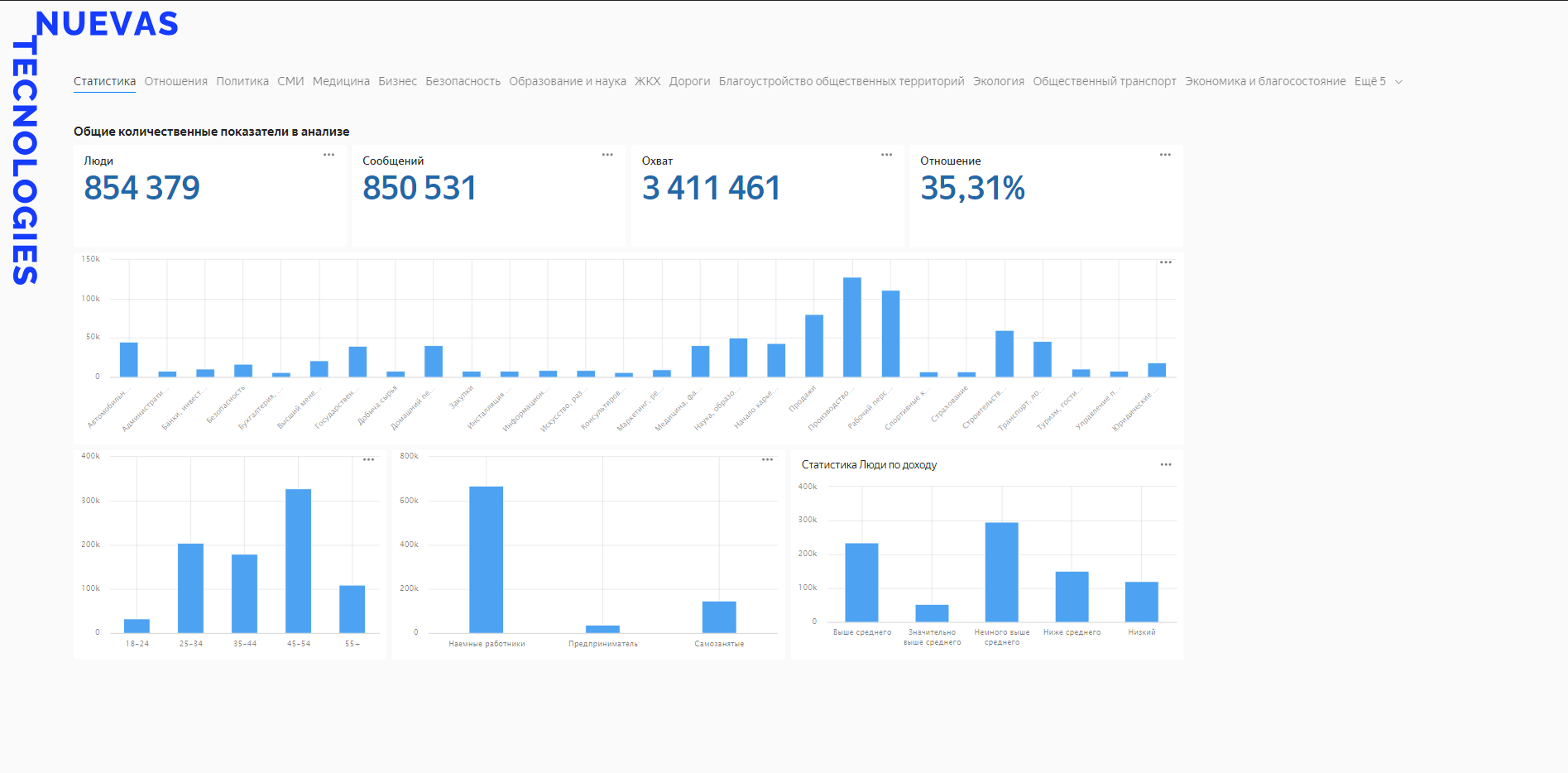


# Интерфейс системы

## Общий вид

В верхней части дэшборда, открывающегося по ссылке после авторизации, располагается перечень доступных вкладок с возможностью выбора каждой из них посредством нажатия левой кнопки мыши.

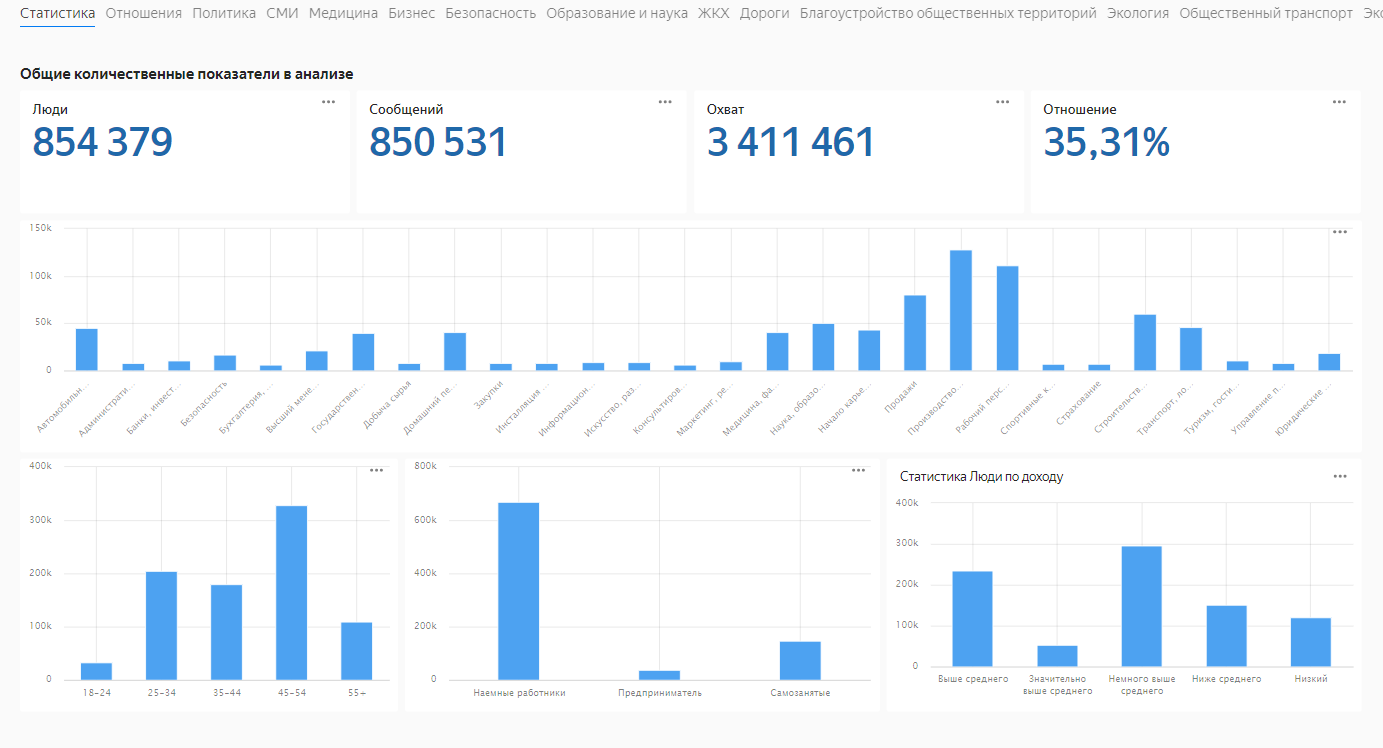
Основная часть дэшборда представлена графиками, диаграммами, тепловыми картами и статистическими таблицами, визуализирующими результаты проведенного Системой анализа и мониторинга. Содержание каждой из вкладок различается в зависимости от предназначения вкладки.



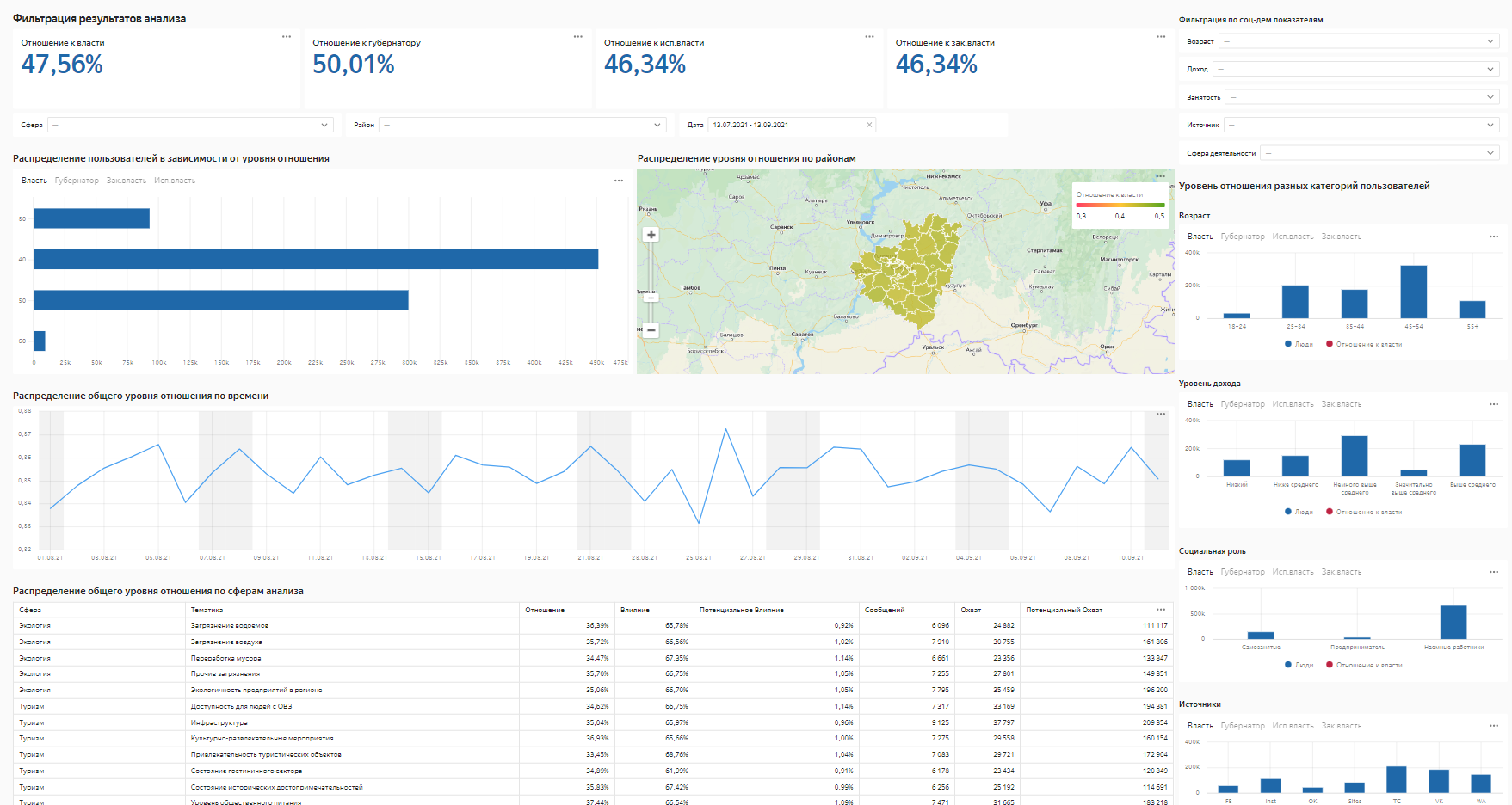
## Перечень основных вкладок

Основными вкладками являются:

* **Статистика** – статистическая информация в части количества уникальных пользователей в анализе, общего количества проанализированных сообщений и охвата. Также на ссылке представлено распределение пользователей в анализе в виде столбчатых диаграмм по уровню дохода, сфере занятости, возрасту и типу занятости



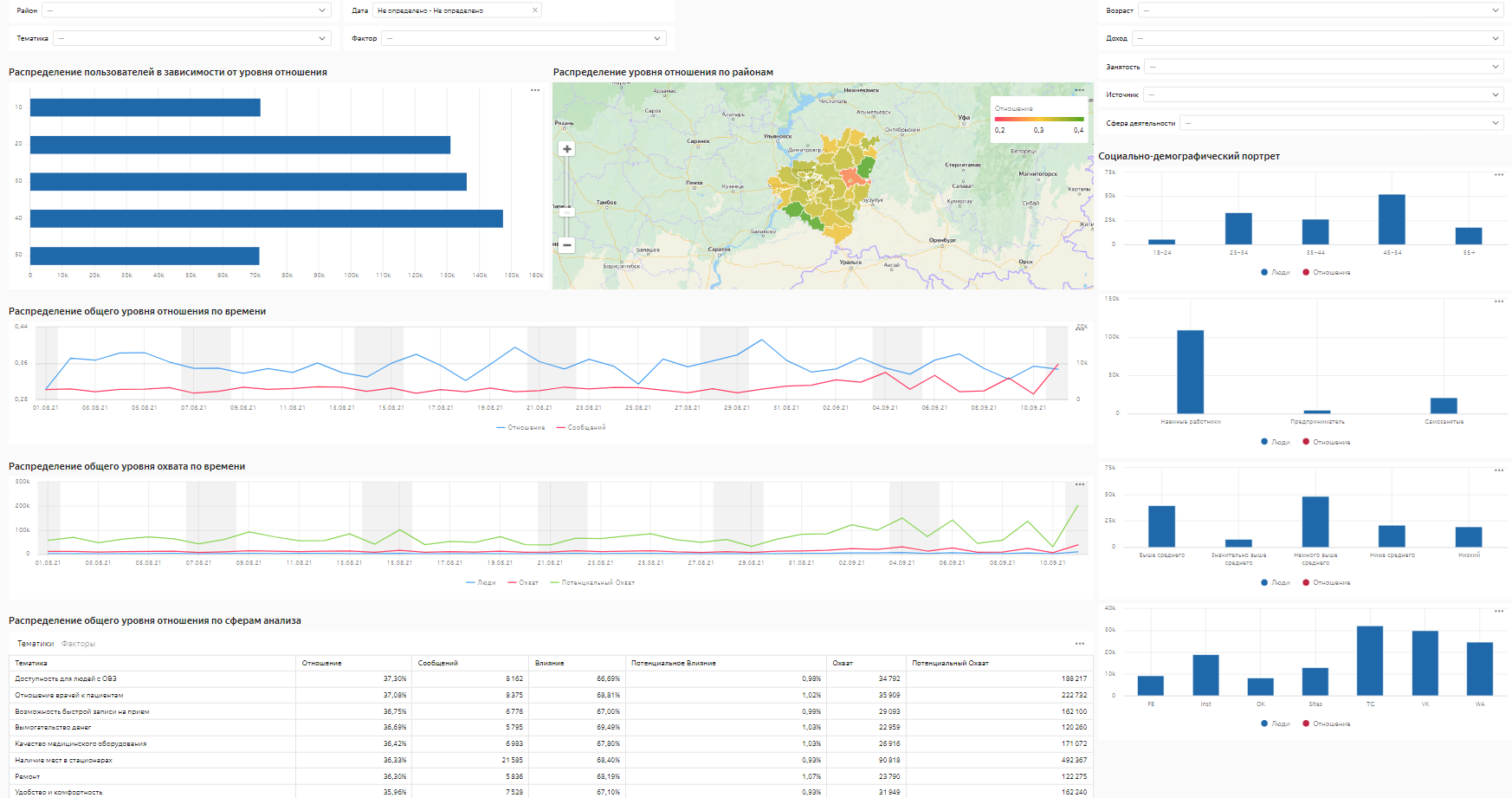
* **Отношение** – основная вкладка с общей аналитической информацией, в рамках которой отображаются данные о совокупном уровне отношения пользователей к различным уровням региональной власти, распределение пользователей по различным категориям, а также основные показатели, демонстрирующие сферы и тематики, из которых складывается совокупное отношение.



Последующие вкладки имеют аналогичный формат и детализируют информацию в части отношения пользователей к той или иной сфере государственного управления и существующих проблемных тематик в ней.

Структурно данные вкладки состоят из различного распределения уровня отношения пользователей в виде диаграмм, графиков и тепловых карт, а также таблицы с общими статистическими показателями по каждой из выявленных тематик в рамках анализируемой сферы.

Также есть возможность фильтрации данных посредством различных взаимосвязанных фильтров.



* *Медицина*
* *Бизнес, предпринимательство*
* *Дороги*
* *Благоустройство общественных территорий*
* *Безопасность, правопорядок, судебная деятельность*
* *Экология*
* *Общественный транспорт*
* *Образование и наука*
* *Экономика и благосостояние*
* *Социальная защита и обеспеченность*
* *Спорт и спортивные мероприятия*
* *Строительство и развитие инфраструктуры*
* *ЖКХ*
* *Молодежная политика*
* *Культура*
* *Общественные мероприятия*
* *Туризм*
* *Политика*
* *Рейтинг источников*

## Перечень основных показателей на вкладках

1. **Отношение** – уровень отношения пользователей к конкретному фактору/проблеме/сфере/тематике. Выражен в процентах и показывает, насколько хорошо или плохо отношение людей. Интерпретировать уровень отношения можно следующим образом:

* 0%-20% - резко негативное отношение (уровень критичности -максимальный, люди кипят)
* 20%-40% - умеренно негативное отношение (проблема есть, люди недовольны, но точки кипения еще не достигли)
* 40%-60% - нейтральное отношение (даже если проблема и существует, четко выраженного негатива не столь много, проблема второго плана. Также в данном случае проблема может быть спорной и часть населения может быть настроена относительно позитивно, что выравнивает картину происходящего)
* 60%-80% - умеренно позитивное отношение (как правило в таком случае проблема себя уже изжила, либо была решена и обсуждение носит преимущественно позитивный характер)
* 80%-100% - резко позитивное отношение (крайне редкая ситуация, характеризующаяся восторженной реакцией со стороны жителей)

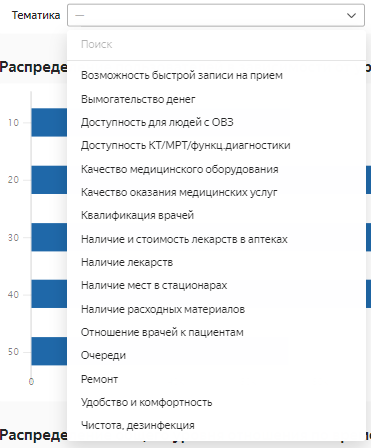
1. **Охват** – количество уникальных пользователей, которые увидели, прочитали, отреагировали или приняли участие в дискуссии
2. **Влияние** – уровень влияния проблематики на людей из числа тех, кто попал в охват. Выражен в процентах и показывает на какой процент пользователей из числа охвата проблематика повлияла негативно (если она негативная) или позитивно (если она позитивная в соответствии с показателем «отношение»).
3. **Сообщения** – кол-во выявленных и проанализированных сообщений, оставленных пользователями
4. **Потенциальное влияние** – влияние проблематики на совокупный уровень отношения пользователей к губернатору. Выражен в процентах и показывает, насколько процентных пунктов поднимется отношение пользователей к губернатору в случае успешного решения данной проблемы
5. **Потенциальный охват** – кол-во уникальных пользователей, которые могут увидеть или которых может затронуть указанная проблематика. Данный показатель является прогнозным и рассчитывается по большей части на основании анализа категорий пользователей, попавших в изначальных охват. Отвечает на вопрос, какое количество пользователей в перспективе может коснуться проблематика в случае, если не приступить к ее оперативному решению.



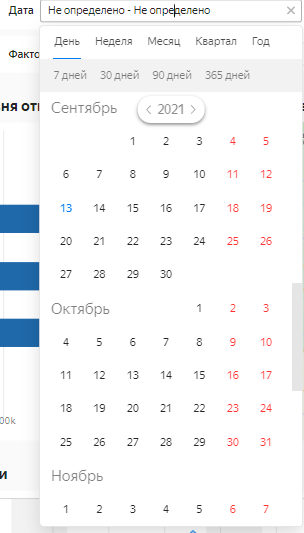
# Принципы сортировки результатов анализа

Каждая из вкладок содержит **фильтры,** при выборе которых динамично изменяется информация, отображаемая на клиентском интерфейсе. **Фильтры бывают двух уровней**:

* **1 уровень:**
* Тематика (возможность выбора определенной тематики в рамках выбранной вкладки)



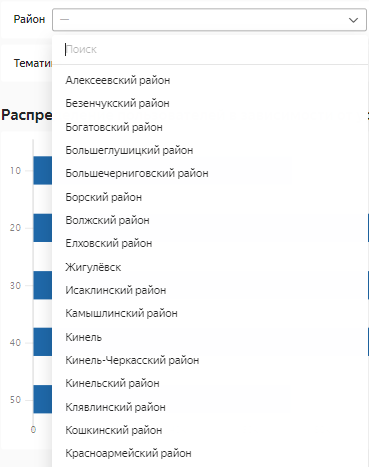
* Период (возможность выбора периода отображения результатов анализа)



* Территориальное деление
* **2 уровень:**
* Сфера деятельности
* Тип занятости
* Уровень дохода
* Возраст
* Источник

## Фильтры по району

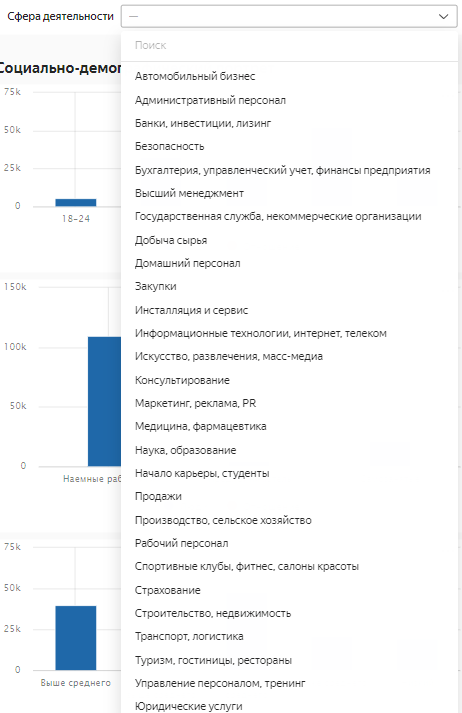
Данные на дашборде могут быть отсортированы по территориальному принципу (по району региона или же по району города при выборе конкретного города). После выбора фильтра все данные на вкладке изменятся в соответствии с выбранным фильтром.



## Фильтры по сфере деятельности

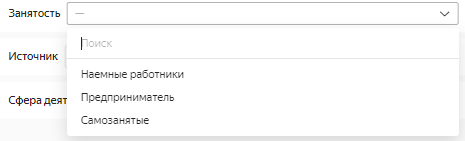
Также фильтрация возможна по сфере деятельности пользователей. При выборе конкретного значения из выпадающего списка данные на дашборде будут отображены только по выбранной категории пользователей. Базово разбивка происходит по следующим категориям:

* *Автомобильный бизнес*
* *Административный персонал*
* *Банки, инвестиции, лизинг*
* *Безопасность*
* *Бухгалтерия, управленческий учет, финансы предприятия*
* *Высший менеджмент*
* *Государственная служба, некоммерческие организации*
* *Добыча сырья*
* *Домашний персонал*
* *Закупки*
* *Инсталляция и сервис*
* *Информационные технологии, интернет, телеком*
* *Искусство, развлечения, масс-медиа*
* *Консультирование*
* *Маркетинг, реклама, PR*
* *Медицина, фармацевтика*
* *Наука, образование*
* *Начало карьеры, студенты*
* *Продажи*
* *Производство, сельское хозяйство*
* *Рабочий персонал*
* *Спортивные клубы, фитнес, салоны красоты*
* *Страхование*
* *Строительство, недвижимость*
* *Транспорт, логистика*
* *Туризм, гостиницы, рестораны*
* *Управление персоналом, тренинг*
* *Юридические услуги*



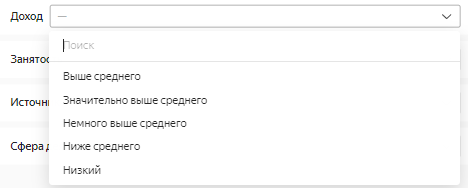
## Фильтры по типу занятости

Фильтрация по типу занятости позволяет выделить ту или иную категорию пользователей. Базово выделяются следующие категории пользователей: наемные работники, самозанятые и предприниматели.



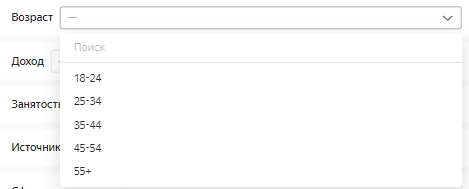
## Фильтры по уровню дохода

Фильтрация по уровню дохода происходит в соответствии с относительной классификацией, основанной на расчете среднего уровня дохода жителей анализируемого региона или города. На дашборде представлена следующая классификация по уровню дохода:



## Фильтры по возрасту

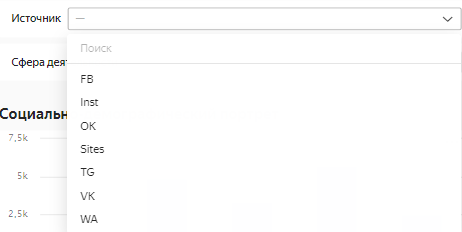
Аналогичным образом происходит и фильтрация пользователей по возрастным категориям. Базово выделяются несколько основных категорий пользователей, однако деление может также осуществляться с шагом вплоть до одного года.



## Фильтры по источнику

Также возможно отсортировать результаты анализа в зависимости от источников сообщений. Базово среди основных источников выделяются:

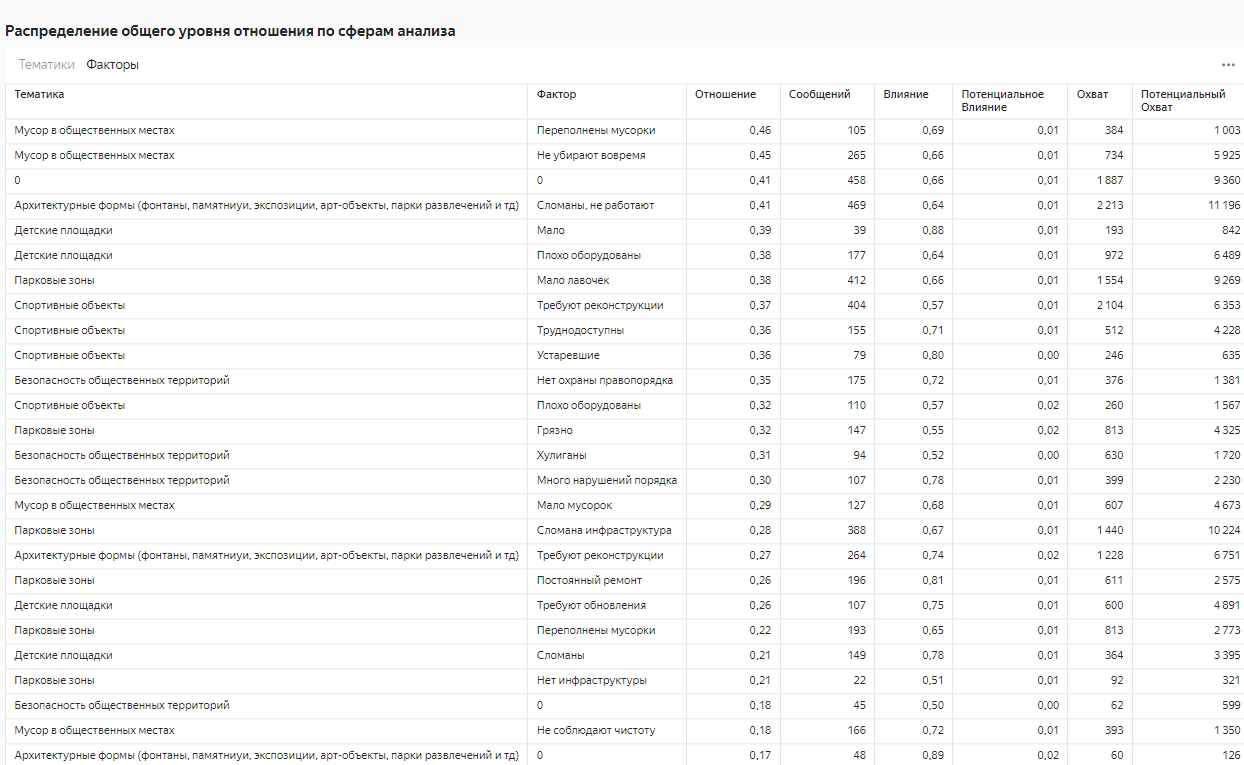
* ВКонтакте
* Фейсбук
* WhatsApp
* Instagram
* Одноклассники
* Telegram
* Прочие сайты и форумы



# Принципы распределения результатов анализа

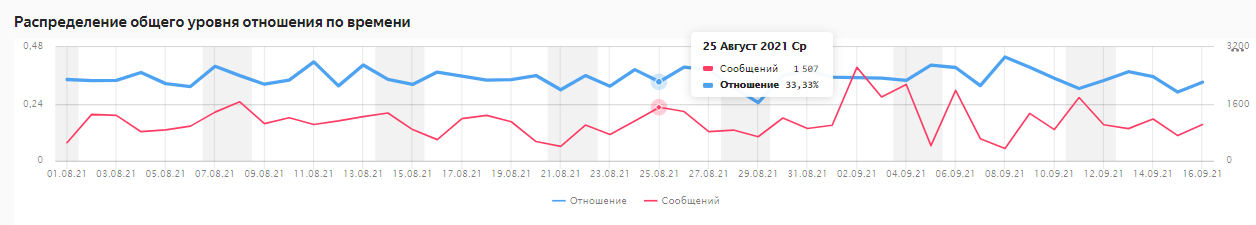
## Распределение данных по тематикам и факторам

В рамках каждой из вкладок в нижней части дашборда располагается таблица с распределение основных показателей по тематикам. Также при выборе вкладки «факторы» отображается распределение каждой из тематик на факторы в качестве следующего уровня углубления и детализации.



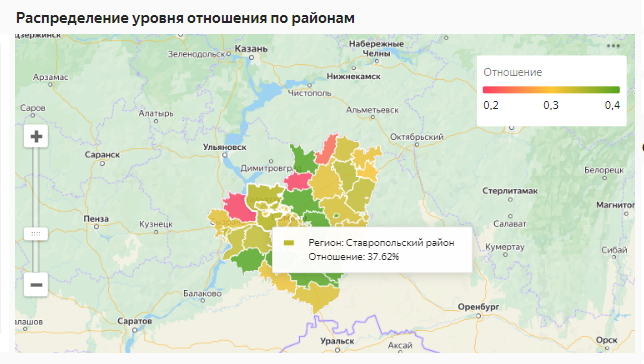
## Распределение уровня отношения и сообщений по времени

Также на дашборде на вкладках отображается линейный график динамики изменения уровня отношения в привязке к количеству людей, участвующих в дискуссии в зависимости от выбранных фильтров. На графике отдельно на каждую дату из анализируемого периода рассчитано конкретное значение.



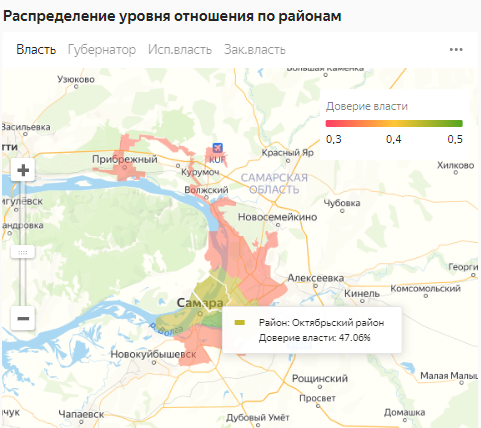
## Распределение уровня отношения по районам региона

Уровень отношения жителей к выбранной сфере, тематике или фактору (в зависимости от выбранных на вкладке фильтров) распределяется по районам анализируемого региона. Визуализация данного процесса реализована посредством построения тепловой карты распределения с нанесенными на карту границами каждого из районов.



## Распределение уровня отношения по районам города

Аналогичным образом при выборе конкретного города, достаточно крупного для того, чтобы имело смысл распределение отношения по районам города. Распределение в данном случае также визуализируется в виде тепловой карты города. При наведении курсором на отдельный район – появляется всплывающее окно с наименованием района и числовым показателем уровня отношения жителей.



## Распределение количество людей, охвата и потенциального охвата сообщений по времени

Динамика изменения количества людей в привязке к охвату и потенциальному охвату отображается на линейном графике в разбивке по дням в рамках анализируемого периода. При наведении на график курсором появляется всплывающее окно с числовыми значениями каждого из показателей.

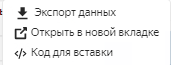


# Экспорт данных

## Экспорт распределения по тематикам и факторам

На каждой вкладке в правом верхнем углу таблицы распределения тематик, факторов и числовых показателей находится иконка

При нажатии на данную иконку появляется всплывающее окно с возможность выбора формата выгружаемых данных и после выбора автоматически происходит процесс экспорта в выбранном виде.



## Экспорт графиков и диаграмм

Аналогичным образом процесс экспорта данных строится и в случае экспорта графиков и диаграмм. На каждом графике/диаграмме в правом верхнем углу таблицы распределения тематик, факторов и числовых показателей находится иконка

При нажатии на данную иконку появляется всплывающее окно с возможность выбора формата выгружаемых данных и после выбора автоматически происходит процесс экспорта в выбранном виде.

