



КРИБРУМ
Мы слушаем сеть

Система визуализации данных из социальных медиа и СМИ «Крибрум. Гео»

Описание Системы

Версия 1.0

АННОТАЦИЯ

Система визуализации данных из социальных медиа и СМИ «Крибрум. Гео» (далее — Система «Крибрум. Гео» или Система) предназначена для отображения ключевых для заказчика показателей по объекту исследования на интерактивной географической карте. Объектом исследования может быть любой процесс, вопрос, явление, проблема в целом, отдельные их аспекты, стороны, отношения между сторонами и целыми изучаемого объекта (социально политический процесс, мероприятие, интересы и т.д.), обсуждение или упоминание которого ведется в сети Интернет.

Назначение для руководителей и пиар-служб:

- Федеральных и региональных органов государственной власти;
- Силowych структур;
- Коммерческих и некоммерческих организаций;
- Государственных предприятий.

Функциональные возможности:

- цветовая индикация выбранного географического объекта (например, страна, макро-регион, регион, район, город и др.) на интерактивной географической карте;
- детализация информации на интерактивной географической карте в соответствии с административно-территориальным делением выбранного географического объекта;
- визуализация ключевых для заказчика показателей (для каждого проекта определяется индивидуальный набор отображаемых показателей) в виде графиков, гистограмм, диаграмм и т.д.:
 - шкала, демонстрирующая уровень присутствия объекта исследования (например, протестная активность, социальная напряженность, тональность, количество публикаций за день в информационном пространстве и другие показатели);
 - рейтинг источников в виде диаграмм;
 - динамика публикаций;
 - соотношения нескольких процессов и явлений;
 - составление прогнозных моделей;
 - топ авторов (поддержка/критика);
 - отображение наиболее обсуждаемых тем;
 - отображение изменения обсуждения и упоминания объекта исследования за выбранный период;
 - отображение отношения пользователей к объекту исследования и т.д.
- просмотр динамики присутствия объекта исследования, с настраиваемой глубиной ретроспективы от 2 недель до 3 лет.

Данный документ содержит описание Системы.

ОГЛАВЛЕНИЕ

1	Общие сведения о системе.....	5
1.1	Наименование Системы	5
1.2	Разработчик Системы	5
1.3	Область применения Системы.....	5
1.4	Программное обеспечение, необходимое для функционирования Системы	5
1.5	Языки программирования, на которых написана программа.....	6
2	Функциональное назначение.....	7
3	Описание технической архитектуры системы.....	8
3.1	Используемые методы сбора и анализа	8
3.2	Алгоритм обработки информации в Системе	8
4	Описание функциональных возможностей Системы	10
5	ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ	12
5.1	Шифрование	12
5.2	Просмотр на мобильных устройствах.....	12

ПЕРЕЧЕНЬ ТЕРМИНОВ И СОКРАЩЕНИЙ

Атрибут	— признак публикации (дата размещения/обнаружения, время размещения/обнаружения, автор, источник, тональность, география, тематика)
География	— географическая принадлежность публикации, которая определяется: - по географически привязанным текстовым маркерам («у нас в Перми» или «рядом с Мариинкой») - по географии автора (на основе информации его профиля)
Информационное пространство	— любое средство распространения информации для широкого (как правило, неограниченного) круга лиц: средство массовой информации, новостной интернет-ресурс (портал), различные виды социальных медиа, общедоступные каналы распространения информации в мессенджерах
Информационный повод (инфоповод)	— одно или несколько событий, которые являются стимулом к информационному обсуждению, формированию мнений и взглядов по теме события
Источник	— То же, что Информационное пространство
Объект исследования	— то, что подвергнуто целенаправленному, систематизированному изучению в информационном пространстве. Объектом исследования может быть любое название компании, персоны, бренда, продукта, услуги, акции, географического объекта, события, явления, процесса и т.п.
Публикация	— информационный блок, опубликованный в источниках информации. Публикацией (сообщением) может быть новость, статья, интервью, репортаж, видеосюжет, аудиозапись, пост, репост, комментарий, твит, видеоролик.
Социальные медиа (соцмедиа)	— источники информации, позволяющие любым пользователям сети Интернет (в том числе пользователям, не являющимися профессиональными авторами) создавать, размещать, распространять и обсуждать публикации
Средство массовой информации, СМИ	— источник, зарегистрированный в порядке, предусмотренном Законом Российской Федерации от 27.12.1991 № 2124-1 «О средствах массовой информации». Сведения о средствах массовой информации содержатся в Перечне наименований зарегистрированных СМИ Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций на интернет-странице https://rkn.gov.ru/mass-communications/reestr/media/
Тональность	— оценочное отношение автора к объекту высказывания. Может принимать значения «положительно» (автор хвалит объект высказывания), «отрицательно» (автор ругает объект высказывания) или «нейтрально» (объект высказывания упомянут без оценки)

1 ОБЩИЕ СВЕДЕНИЯ О СИСТЕМЕ

1.1 Наименование Системы

Наименование: «Крибрум. Гео»

Обозначение: Система визуализации данных из социальных медиа и СМИ «Крибрум. Гео»

1.2 Разработчик Системы

Разработчиком Системы является АО «Крибрум»

1.3 Область применения Системы

Система «Крибрум. Гео» предназначена для отображения на интерактивной карте присутствия объекта исследования в информационном пространстве.

1.4 Программное обеспечение, необходимое для функционирования Системы

Система «Крибрум. Гео» распространяется в виде интернет-сервиса, установка на стороне Заказчика не требуется.

Программное обеспечение компьютеров, на которых устанавливается Система, должно отвечать минимальным требованиям, указанным ниже.

Системное и базовое ПО, которое должно работать на сервере подсистемы интеграции с мастер-системой:

- ОС Oracle Linux 8;
- БД: PostgreSQL 13 и выше;
- Веб-серверы: nginx 1.21 + apache 2.4.50/mod_perl и выше;
- Интерпретатор: PHP 7.0, Perl 5.30.

Системное и базовое ПО типового автоматического рабочего места (АРМ), используемого для доступа к Системе:

АРМ представляет собой клиентское приложение (веб-интерфейс). Для работы АРМ необходима одна из следующих операционных систем:

- Windows 7, Windows 8, Windows 10, Windows 11;
- Mac OS X 10.6 или более поздней версии.

Для работы также необходимо наличие веб-браузера:

- Opera версии 88 или выше;
- Google Chrome версии 95.0 или выше;
- Mozilla Firefox версии 90 или выше;
- Safari версии 15.2 или выше.

1.5 Языки программирования, на которых написана программа

Языки программирования:

- SQL;
- Perl5;
- PHP7;
- HTML;
- CSS;
- JavaScript.

2 ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ

Система выполняет следующие функции:

- отбирает публикации по объекту исследования за определенный период времени;
- определяет и присваивает географию публикации;
- подсчитывает требуемые показатели (например, протестная активность, социальная напряженность, тональность, количество публикаций и проч.) по разным атрибутам (например, регионам, авторам, источникам и т.д.);
- строит графики и диаграммы;
- отображает ключевые данные и показатели по объекту исследования в информационном пространстве на интерактивной географической карте;
- предоставляет доступ к ретроспективным данным объекта исследования, глубина ретроспективы от 2 недель до 3 лет.

Полученные данные отображаются:

- на интерактивной географической карте:
 - цветовая индикация в зависимости от одного или нескольких показателей (например, количество публикаций, тональность, протестная активность, социальная напряженность и другие показатели);
 - детализация информации в соответствии с административно-территориальным делением выбранного географического объекта (например, страна, макро-регион, регион, район, город и др.);
- на одной или нескольких панелях «Информационная картина»:
 - шкала, демонстрирующая уровень присутствия объекта исследования (например, протестная активность/тональность и другие показатели);
 - рейтинг источников в виде графиков;
 - индивидуальный набор визуализируемых показателей исследования в виде графиков, диаграмм и т.д., например:
 - динамика публикаций;
 - соотношения нескольких процессов и явлений;
 - составление прогнозных моделей;
 - топ авторов (поддержка/критика);
 - наиболее обсуждаемые темы;
 - отношение пользователей к объекту;
 - другие показатели, необходимые для оперативного отображения.

3 ОПИСАНИЕ ТЕХНИЧЕСКОЙ АРХИТЕКТУРЫ СИСТЕМЫ

3.1 Используемые методы сбора и анализа

Объем, характер и разнообразие данных, которые необходимо получить и визуализировать, требуют применения специализированных методов и создания специализированных программных средств.

К методам поиска и хранения данных относятся:

- спайдеры (программа, которая выполняет просмотр страниц Интернета и собирает разнородную информацию) для индексирования информационного пространства социальных медиа и СМИ;
- «поточный» поиск для уменьшения используемых технических средств и улучшения скорости поиска;
- специализированные хранилища для хранения и оперативного доступа к большому объему информации.

К методам анализа и визуализации данных относятся:

- составление поисковых алгоритмов, обеспечивающих полноту и чистоту (отсутствие информационного мусора) при сборе данных по заданному объекту мониторинга;
- автоматическое распознавание текстов на изображениях;
- построение диаграмм с количеством публикаций;
- разработка формул подсчета показателей;
- составление алгоритмов, обеспечивающих цветовую индикацию рассчитанного показателя на интерактивной географической карте.

К методам передачи данных, полученных в результате поиска и анализа, из хранилищ в интерфейс Системы используется API.

3.2 Алгоритм обработки информации в Системе

Система «Крибрум. Гео» отбирает публикации, относящиеся к объекту исследования для выбранного географического объекта и глубины сохранения ретроспективы исследования, анализирует их и определяет атрибуты, подсчитывает показатели и визуализирует полученные данные. А именно:

Собирает разнородную по типам информацию из социальных медиа и СМИ и сохраняет ее в базах данных:

- тексты и изображения (в том числе изображения с текстами) из публикаций (комментариев, постов, новостей, заметок, статей, рецензий и т.д.);
- атрибуты публикаций (дата и время размещения и обнаружения, автор, источник и т.д.);
- индивидуальные действия (пост, репост, лайк, комментариев и т.д.).

Отбирает данные (на основе разработанной лингвистической модели), относящиеся к объекту исследования, соответствующие выбранному географическому объекту и определенному периоду сохранения ретроспективы.

Обрабатывает массивы данных, относящихся к объекту исследования:

- определяет и присваивает атрибуты публикаций (географию, тональность, дату и время размещения и обнаружения, автора, ссылку источник и т.д.);

- подсчитывает показатели (протестная активность, социальная напряженность, индекс медиаприсутствия, количество публикаций по источникам, количество публикаций по тональности и другие показатели).

Визуализирует полученные данные в графическом интерфейсе:

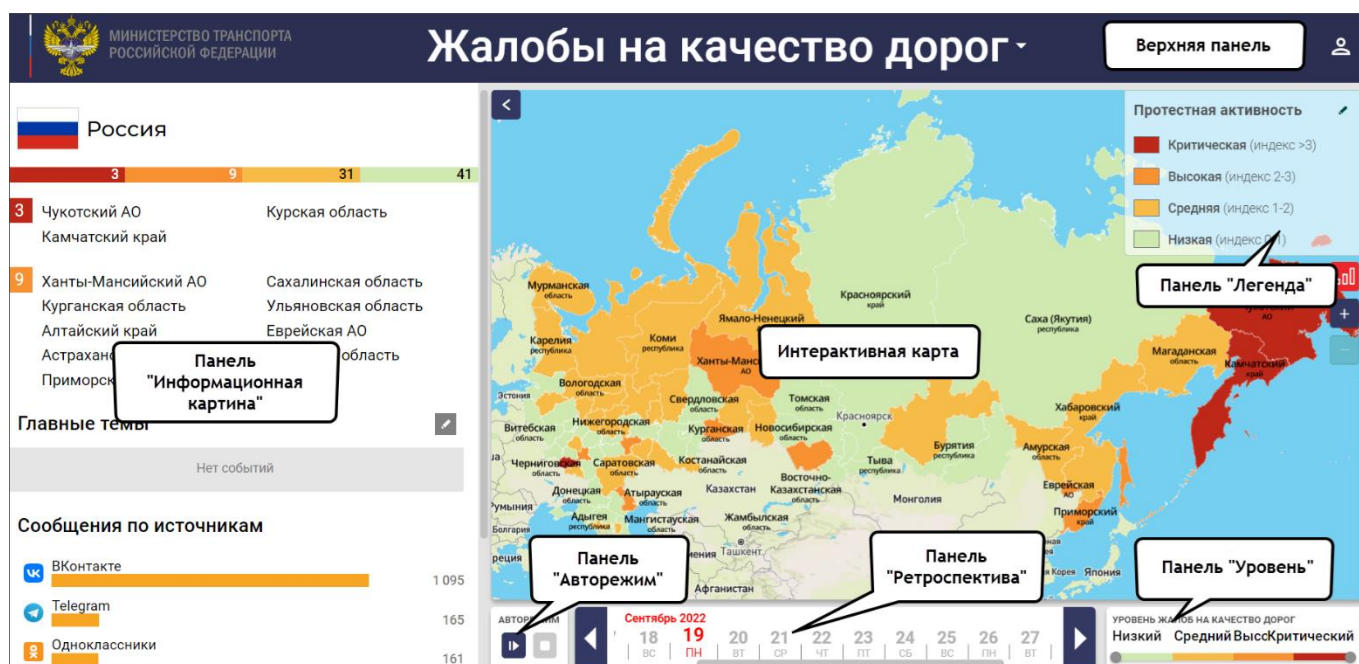
- цветовая индикация на интерактивной географической карте выбранной территории (например: страна, макро-регион, регион, город) с детализацией следующего уровня (например: регион, город, район) в зависимости от показателя, определенного заказчиком (цветом могут выделяться, в частности, протестная активность, социальная напряженность, количественные и другие показатели);
- шкала, демонстрирующая уровень присутствия объекта исследования (например, протестная активность, тональность и другие показатели);
- рейтинг источников в виде диаграммы;
- другие показатели.

4 ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ВОЗМОЖНОСТЕЙ СИСТЕМЫ

Вход в Систему осуществляется по URL-адресу <https://tabula.kribrum.ru/> с авторизацией по логину и паролю.

Система «Крибрум. Гео» состоит из одного окна, на котором размещены:

- интерактивная географическая карта, где цветовой индикацией отображается присутствие одного или нескольких показателей (например, протестная активность, тональность, количество публикаций и другие показатели) объекта исследования в информационном пространстве;
- панели различной с информацией по интересующим заказчика показателям в виде диаграмм, аналитических справок и т.д. (например, главные события, количество публикаций и др.)



Цветовая гамма и количество уровней показателей объекта исследования настраиваются по требованию Заказчика. Показатели, которые будут визуализированы, определяются Заказчиком под конкретный проект.

Данные по объекту исследования отображаются за установленный заказчиком период времени (по умолчанию отображаются данные за текущий день). Для просмотра ретроспективных данных можно выбрать любой день на панели «Ретроспектива».

Система поддерживает настройку отображения только определенных уровней показателей объекта мониторинга (например, среднего и/или высокого) на панели «Уровень».

Система может автоматически перелистывать ретроспективные данные на интерактивной карте (слайды) за выбранный заказчиком период сохранения информации (от 2 недель до 3 лет) (авторежим).

Панель «Легенда» предназначена для отображения и редактирования значений диапазонов выбранного показателя для каждого уровня.

Панель «Информационная картина» содержит следующую информацию либо по выбранному географическому объекту (мир/страна/регион/район/город), либо по территориальной единице (страна/регион/район/город/муниципальное образование), входящей в географический объект:

Информация по географическому объекту	Информация по территориальной единице
<ol style="list-style-type: none"> 1. название выбранного географического объекта (страна/регион/район/город); 2. шкала уровней показателя с количеством территориальных единиц, соответствующих каждому уровню; 3. список (списки) территориальных единиц, соответствующий выбранному (соответствующих выбранным) заказчиком уровню (уровням). 4. наиболее обсуждаемые темы в выбранном географическом объекте; 5. рейтинг источников с количеством публикаций. 	<ol style="list-style-type: none"> 1. название выбранной территориальной единицы географического объекта (регион/район/город/муниципальное образование). 2. место в рейтинге по показателю; 3. наиболее обсуждаемые темы в территориальной единице; 4. рейтинг источников с количеством публикаций

5 ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

5.1 Шифрование

Передача данных от сервера Систему к клиентскому рабочему месту является защищённой. Для этого используется криптографический протокол TLS обеспечивающий:

- шифрование информации, передаваемой от одного компьютера к другому;
- аутентификацию авторства передаваемой информации;
- целостность информации.

Наш сертификат подтвержден в цепочке официальных центров сертификации.

Протокол TLS осуществляет защиту данных пользователя от перехвата третьими лицами, предотвращает несанкционированный доступ и возможную прослушку трафика. Он обеспечивает защиту данных пользователя от момента аутентификации до окончания сессии.

5.2 Просмотр на мобильных устройствах

Просмотр функционала, доступного в веб-версии Системы, также возможен на мобильных IOS и Android устройствах.