

ПЛАТФОРМА МОНИТОРИНГА И АНАЛИЗА СОЦИАЛЬНЫХ МЕДИА КРИБРУМ

04.03.2021

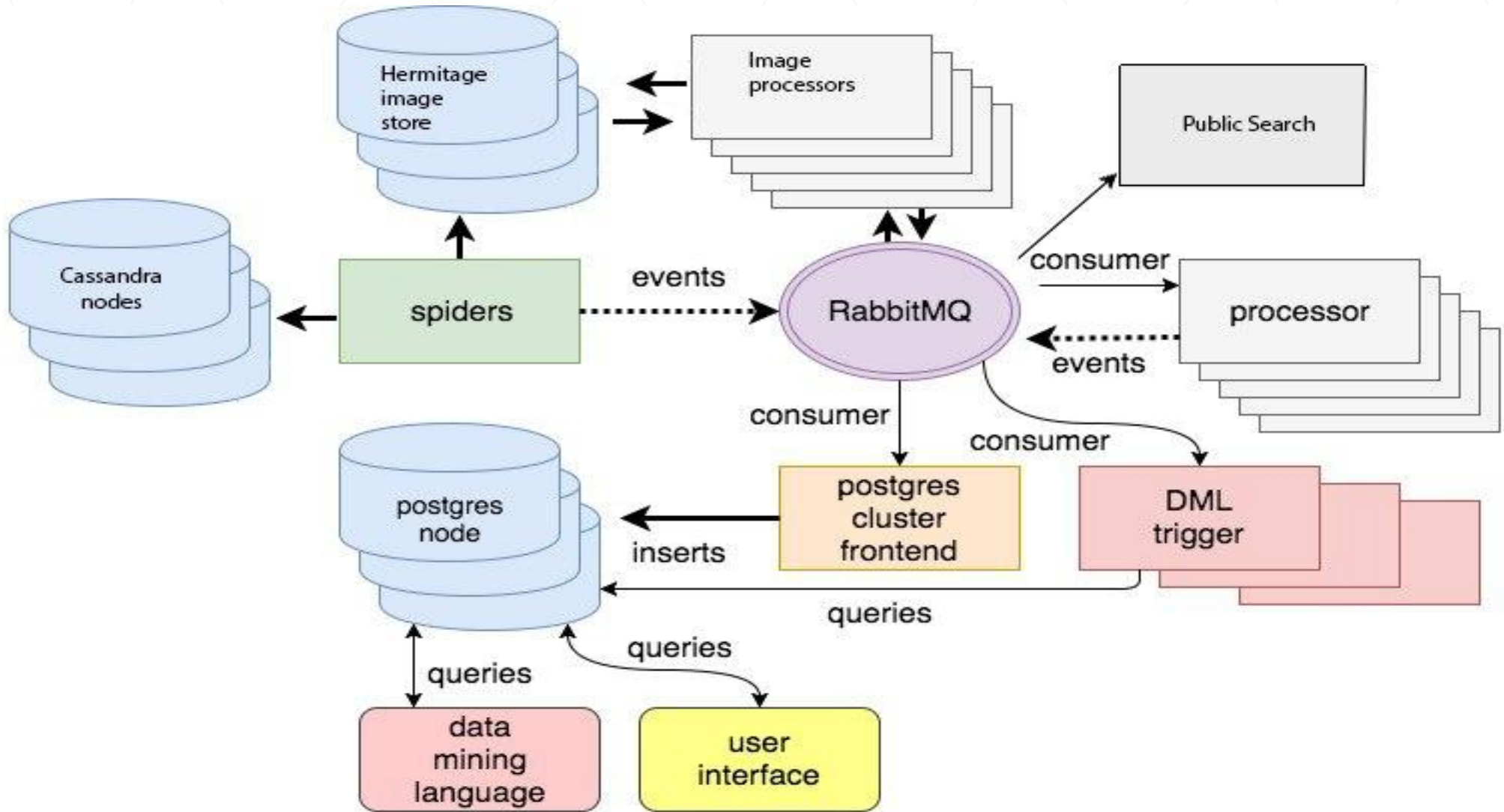
Бронзинский М.Д.



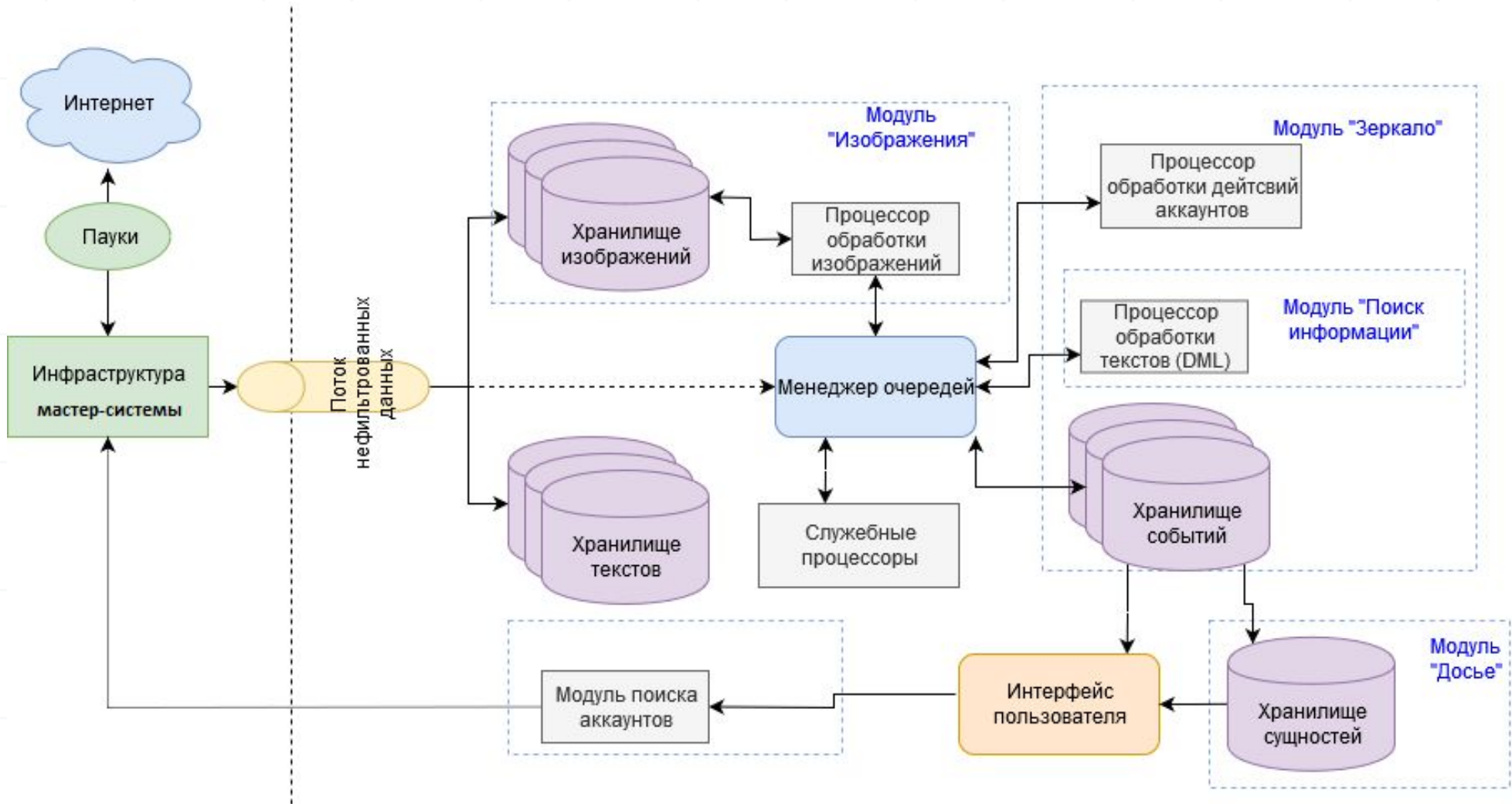
КРИБРУМ

Мы слушаем сеть

Архитектура платформы Крибрум



Размещение в контуре Заказчика





ПЛАН

- Социальные сети, основные понятия, особенности сбора.
- Общие сведения об архитектуре платформы, взаимодействие основных компонент
- Обзор систем Платформы Крибрум

Социальные сети, основные понятия, особенности сбора

04.03.2021



КРИБРУМ

Мы слушаем сеть

Основные термины

Автор – это пользователь от имени, которого, опубликовано сообщение на интернет ресурсе (площадке мониторинга);

Источник сообщения – это площадка мониторинга, на которой обнаружено релевантное сообщение по отношению к объекту мониторинга;

Оригинал - сообщение, опубликованное первым в ряду одинаковых или частично повторяющихся;

Перепечатки - ряд одинаковых или частично повторяющихся друг друга сообщений;

Основные термины

Площадки мониторинга – это веб-сайты, которые включены в мониторинг, то есть отслеживаются Системой на предмет появления там сообщений, содержащих упоминания объектов мониторинга;

Система – система автоматизированного сбора и анализа социальных медиа;

Сообщение – это отдельная текстовая публикация в Интернете, содержащая осмысленное упоминание объекта мониторинга на одной из площадок мониторинга. Сообщение может иметь вид записи, комментария, поста, новости, заметки, статьи, рецензии в блоге, микроблоге, социальной сети, форуме, онлайн-СМИ, интернет-магазине или другом виде информационных Интернет-ресурсов и социальных медиа;

Спайдер - программа, часть поисковой системы. Она перебирает страницы Интернета и собирает информацию о них в виде



Основные термины

Объект – это уникально поименованные персона, событие, название компании, название услуги, проекта (акции), географического объекта, природного явления. В понятие «объект» входят все синонимы и виды написания.

Примеры:

- о «Ростелеком, Rostelecom, ростелек, Ростелик» и т.д.;
- о «Путин, Президент, Владимир Владимирович», и т.д.;
- о «Самара»;
- о «Митинг 3 июня.

Основные термины

Группа объектов – совокупность объектов одного структурного уровня, относящихся к одной сфере или тематике.

Примеры:

- о Банки (Сбербанк, ВТБ, Открытие, ПСБ и т.д.);
- о Страховые компании (Росгосстрах, Альфа Страхование и т.д.);
- о Преподаватели ВШЭ (Иванов, Петрова и т.д.);
- о Митинги 2020 (Митинг 3 августа, митинг 15 сентября и т.д.);
- о Проекты Яндекс (Яндекс.Лавка, Яндекс.Такси и т.д.).

Основные термины

Тема – определенная сфера жизнедеятельности общества, государства, компании, социальной группы. Для мониторинга темы разрабатывается специализированная лингвосемантическая модель, а также архитектура проекта, включающая десятки/сотни объектов, тегов, категорий.

Примеры:

- о ЖКХ (тарифы, льготы, изменения в законодательстве и т.д.);
- о Медицина (медицинские учреждения, проекты, новые медицинские услуги и т.д.);
- о Высшее образование (вузы, льготы, руководители учебных заведений, государственная поддержка, отзывы об общежитиях и т.д.);

Основные термины

Направление – совокупность социальных проявлений человеческой активности, имеющих определенный характер и преследующих заранее заданные цели. Для изучения направления привлекаются эксперты, разрабатываются методики анализа инфополя, затем лингвосемантические модели по темам, составляющим направление.

Примеры:

- о Экстремизм (националисты, антифа, исламский фундаментализм, секты и т.д.);
- о Протестная активность (протесты, пикеты, лидеры протестного движения, виды протестов по темам и т.д.);
- о Деструктив в молодежной среде (девиантные молодежные сообщества, субкультуры, асоциальные и деструктивные практики,

ПЛАТФОРМА КРИБРУМ

Крибрум – это **Платформа** потокового многофакторного анализа социальных медиа в режиме реального времени.

На базе Платформы Крибрум разработано нескольких систем и модулей.
Системы Платформы Крибрум:



ПЛАТФОРМА КРИБРУМ

Мониторим все виды социальных медиа

140 млн. текстовых сообщений
60 млн. изображений
500 млн. действий

КАЖДЫЙ
ДЕНЬ

**ВСЕ
СОЦСЕТ
И**
FB, VK,
Instagram,
TikTok и др.

31 000
**ИНТЕРНЕ
Т-СМИ**

**МИКРО-
БЛОГИ**
Twitter и
др.

КАНАЛЫ
Telegram

**БЛОГИ,
ФОРУМЫ**

+
подключае
м
ИСТОЧНИКИ
по запросу

**САЙТЫ
DEEP WEB**
(Tor)

**ВИДЕО-
ХОСТИН
ГИ** YouTube
и др.

- Храним всю информацию с **2014** года
- Объём ретроспективной базы данных — 3 петабайта

ПЛАТФОРМА КРИБРУМ

- Анализируем все виды контента: посты, комментарии, ссылки, шеры, лайки, теги, подписи и т.д.
- Разрабатываем **экспертные модели для анализа данных и прогнозирования** событий в реальной жизни.
- Применяем весь спектр **современных собственных технологий**:
 - искусственный интеллект,
 - машинное обучение,
 - нейронные сети,
 - поисковые технологии,
 - лингвистические технологии,
 - анализ больших данных.

ПЛАТФОРМА КРИБРУМ



- ✔ Модуль оценки тональности
- ✔ Модуль идентификации лиц
- ✔ Модуль прямого поиска
- ✔ Модуль обратного поиска
- ✔ Автоматическая классификация текстов
- ✔ Модуль классификации изображений
- ✔ Модуль детекции рисков

АНАЛИТИЧЕСКИЙ ЦЕНТР КРИБРУМ



- Осуществляет анализ тенденций, рисков, новых угроз, социальной температуры и общественного мнения на основе полученных данных
- Разрабатывает модели для прогнозирования событий в реальной жизни: митинги, протесты, кампании в СМИ, распространение поддельных новостей
- Применяет более 500 сложных социально-психологических моделей для анализа пользователей социальных сетей и прогнозирования реальных событий

• Объединяет программистов

Общие сведения об архитектуре платформы, взаимодействие основных компонент

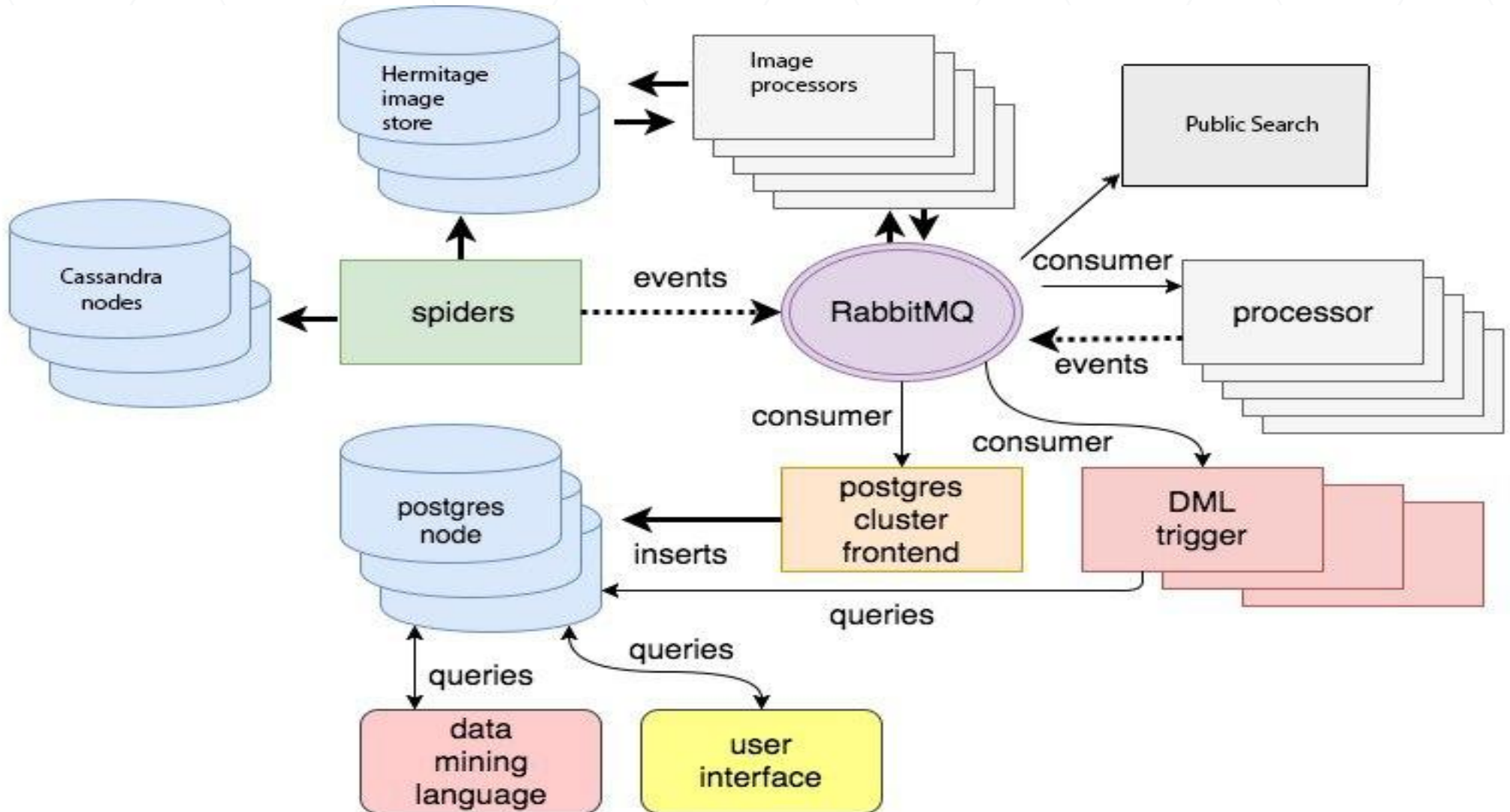
04.03.2021



КРИБРУМ

Мы слушаем сеть

Архитектура платформы Крибрум



Определение событий и типов событий

Событие представляет собой некоторое произошедшее событие, привязанное к конкретному аккаунту внутри системы Крибрум.

События имеют следующие свойства:

- Событие максимально атомарно. Если возможно логически разделить некоторое событие на несколько отдельных, то либо оно должно быть разделено, либо эти более атомарные события уже должны физически существовать.
- Событие всегда иммутабельно. Раз попавшее в систему событие не бывает изменено, а событие, записанное в кластер, остаётся там навечно в оригинальном виде (не удаляется и не модифицируется).



Примеры событий

[event::post](#)

[event::comment](#)

[event::share](#)

[event::profile field work](#)

[event::like](#)

[event::follow](#)

Структура событий

(account_id, timestamp, event)

account_id — ext_id автора в системе:
например, 216824745024532@facebook.com.

publication_time — когда произошло событие
вне системы.

create_time — время создания события
внутри системы.

event — одно из конкретных событий.

Data Mining Language (**DML**)

DML представляет собой один или более шаблонов, которые могут быть произвольно сгруппированы между собой.

Шаблон представляет собой:

- Заданное описание структуры некоторого события
- Опциональное ограничение: булева формула, в которой участвуют только связанные переменные, логические операторы и простые функции.

Шаблоны могут быть сгруппированы одним из следующих двух вариантов.

2. Группой, в которой для срабатывания правила всем поданным на вход событиям должны быть успешно сопоставлены все шаблоны в группе (группировка по "**И**")
3. Группой, в которой для срабатывания правила должен быть успешно сопоставлен хотя бы один шаблон (группировка по "**ИЛИ**")

Группировки могут быть вложенными.

Схема организации полнотекстового поиска





Fort-Morgan t4 Search Engine

Язык поисковых запросов:

word1 word2 – слова запроса в любой словоформе в любом порядке на небольшом расстоянии друг от друга

“word1 word2” – фиксируются словоформы и порядок слов

(word1 word2)/2 – расстояние не больше 2 слов

(word1 word2)! – фиксируется порядок слов

(word1 word2)/* - любое расстояние

word1 (word2 | word3) – альтернатива

word1 –word2 - отрицание

(word1 word2) –((word3 word4)/4 | “word5”) – комбинирование

Особенности:

- Полнотекстовый и атрибутивный поиск
- Поддержка английского, китайского и других языков
- Индексация в реальном времени
- Линейное масштабирование
- Продвинутый язык поисковых запросов

Обзор систем Платформы Крибрум

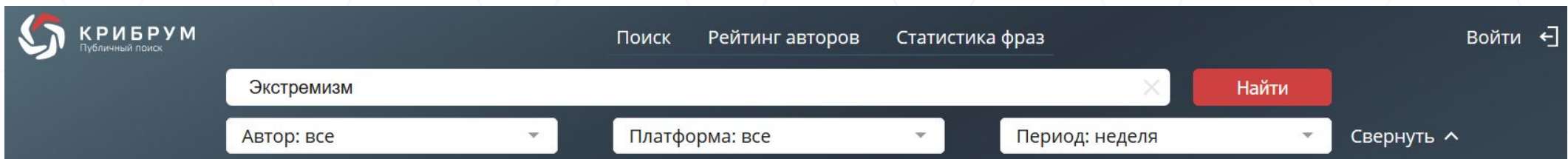
04.03.2021



КРИБРУМ
Мы слушаем сеть

«Крибрум. Публичный поиск»

Предназначена для поиска информации внутри текстов, написанных в социальных медиа.



The screenshot shows the Kribrum search interface. At the top left is the logo and name 'КРИБРУМ Публичный поиск'. Navigation links include 'Поиск', 'Рейтинг авторов', and 'Статистика фраз'. A 'Войти' button is in the top right. The search bar contains the text 'Экстремизм'. Below the search bar are three filters: 'Автор: все', 'Платформа: все', and 'Период: неделя'. A red 'Найти' button is to the right of the search bar. A 'Свернуть' button is at the bottom right of the filter area.

Найдено 3 тысячи записей

Сортировать: **сначала новые**



1

@Александр Полозов ⌚ 6 минут назад

В Сибири к заявителю митинга против «обнуления» пришли за свастику в ...

... был задержан сотрудниками центра по противодействию **экстремизму** и доставлен в полицию. Там на ... просмотра. Это сотрудники центра по противодействию **экстремизму** посчитали достаточным для того, чтобы обвинить ...

www.znak.com/2020-03-16/v_sibiri_k_zayavitelyu_mitinga_p... Подробнее 👁



Helios7 🌞 @geleyos ⌚ 13 минут назад

«МВД: Привлеките активиста Тимура Булатова к ответственности за ...

«МВД: Привлеките активиста Тимура Булатова к ответственности за преследования и **экстремизм.**» - Подпишите петицию! [t.co ChangeOrgRus](https://t.co/ChangeOrgRus)から

twitter.com/geleyos/status/1239499762140938240 Подробнее 👁

«Крибрум. Публичный поиск»

Реализует функции:

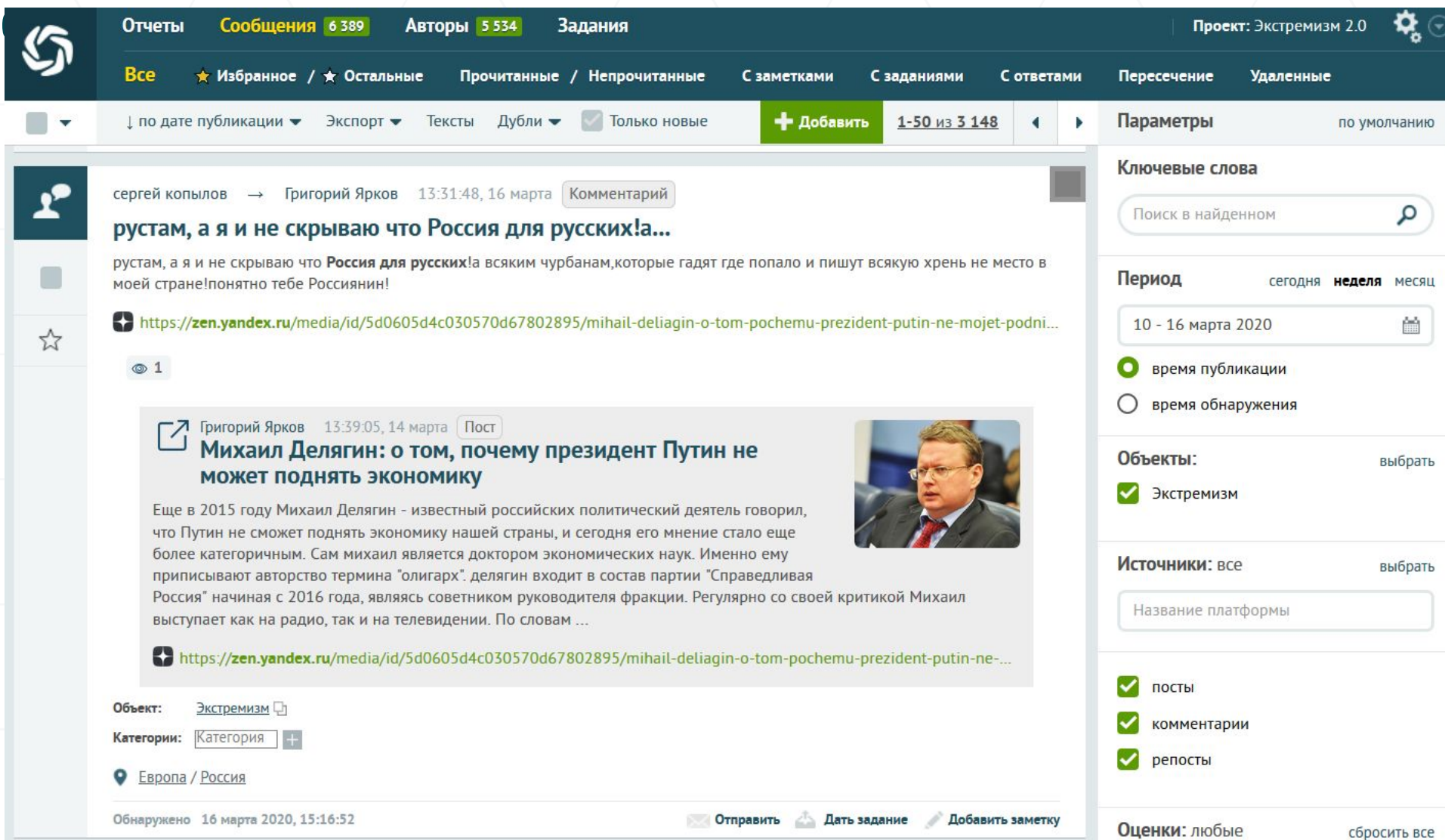
- сбора информации с сайтов и платформ социальных сетей в сети Интернет;
- индексирования, обеспечивающую быстрый поиск по накопленной информации.

Позволяют легко находить нужную информацию в социальных медиа.

«Крибрум. Объекты»

Предназначена для постоянного оперативного сбора и анализа текстовых

соо



Отчеты **Сообщения** 6 389 **Авторы** 5 534 **Задания** Проект: Экстремизм 2.0

Все ★ Избранное / ★ Остальные **Прочитанные** / **Непрочитанные** **С заметками** **С заданиями** **С ответами** **Пересечение** **Удаленные**

↓ по дате публикации Экспорт Тексты Дубли Только новые **+ Добавить** 1-50 из 3 148 **Параметры** по умолчанию

сергей копылов → Григорий Ярков 13:31:48, 16 марта **Комментарий**

рустам, а я и не скрываю что Россия для русских!а...

рустам, а я и не скрываю что **Россия для русских!а** всяким чурбанам,которые гадят где попало и пишут всякую хрень не место в моей стране!понятно тебе Россиянин!

<https://zen.yandex.ru/media/id/5d0605d4c030570d67802895/mihail-deliagin-o-tom-pochemu-prezident-putin-ne-mojet-podni...>

1

Григорий Ярков 13:39:05, 14 марта **Пост**

Михаил Делягин: о том, почему президент Путин не может поднять экономику

Еще в 2015 году Михаил Делягин - известный российских политический деятель говорил, что Путин не сможет поднять экономику нашей страны, и сегодня его мнение стало еще более категоричным. Сам михаил является доктором экономических наук. Именно ему приписывают авторство термина "олигарх". делягин входит в состав партии "Справедливая Россия" начиная с 2016 года, являясь советником руководителя фракции. Регулярно со своей критикой Михаил выступает как на радио, так и на телевидении. По словам ...

<https://zen.yandex.ru/media/id/5d0605d4c030570d67802895/mihail-deliagin-o-tom-pochemu-prezident-putin-ne-...>

Объект: **Экстремизм**

Категории: Категория +

Локация: Европа / Россия

Обнаружено 16 марта 2020, 15:16:52 Отправить Дать задание Добавить заметку

Ключевые слова
Поиск в найденном

Период сегодня **неделя** месяц
10 - 16 марта 2020

время публикации
 время обнаружения

Объекты: выбрать
 Экстремизм

Источники: все выбрать
Название платформы

посты
 комментарии
 репосты

Оценки: любые сбросить все

«Крибрум. Объекты»

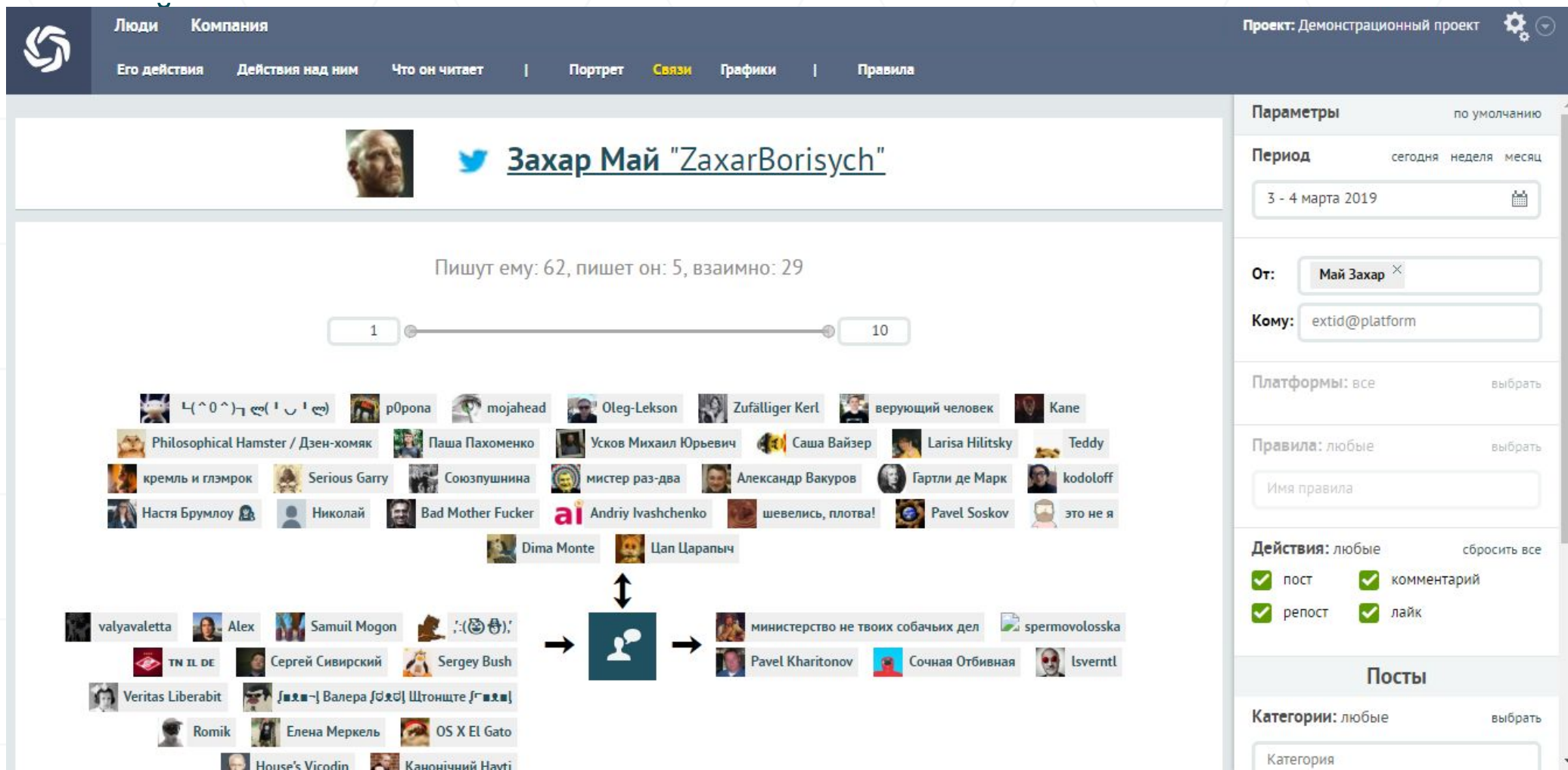
Реализует функции:

- мониторинг социальных медиа;
- сбор, хранение и обработка сообщений;
- автоматическое определение тональности;
- присвоение сообщению тематики, категорий, тегов, географии, Т.Д.;
- составление графиков и диаграмм для дальнейшего анализа;
- распознавание первоисточников и дублей.

Позволяют проводить анализ сообщений по объекту поиска, определять эмоциональное отношение к нему, сортировать полученные данные наиболее оптимальным образом.

«Крибрум. Зеркало»

Предназначена для автоматизированного анализа действий, связей и признаков риска в поведении аккаунтов пользователей социальных



The screenshot displays the 'Kribrum Mirror' interface for a user profile. At the top, there are navigation tabs: 'Люди' (People) and 'Компания' (Company). Below these are sub-tabs: 'Его действия' (His actions), 'Действия над ним' (Actions on him), 'Что он читает' (What he reads), 'Портрет' (Portrait), 'Связи' (Connections), 'Графики' (Charts), and 'Правила' (Rules). The profile header shows a profile picture and the name 'Захар Май "ZaxarBorisych"'. Below the header, statistics are shown: 'Пишут ему: 62, пишет он: 5, взаимно: 29'. A slider below the statistics allows filtering between 1 and 10 items. The main area displays a network graph of associated accounts, with a central node representing the user and arrows pointing to various other accounts. The right sidebar contains 'Параметры' (Parameters) with settings for 'Период' (Period), 'От' (From), 'Кому' (To), 'Платформы' (Platforms), 'Правила' (Rules), 'Действия' (Actions), and 'Посты' (Posts).

Параметры по умолчанию

Период сегодня неделя месяц
3 - 4 марта 2019

От: Май Захар

Кому: extid@platform

Платформы: все

Правила: любые

Действия: любые

- пост
- комментарий
- репост
- лайк

Посты

Категории: любые

«Крибрум. Зеркало»

Реализует следующие функции:

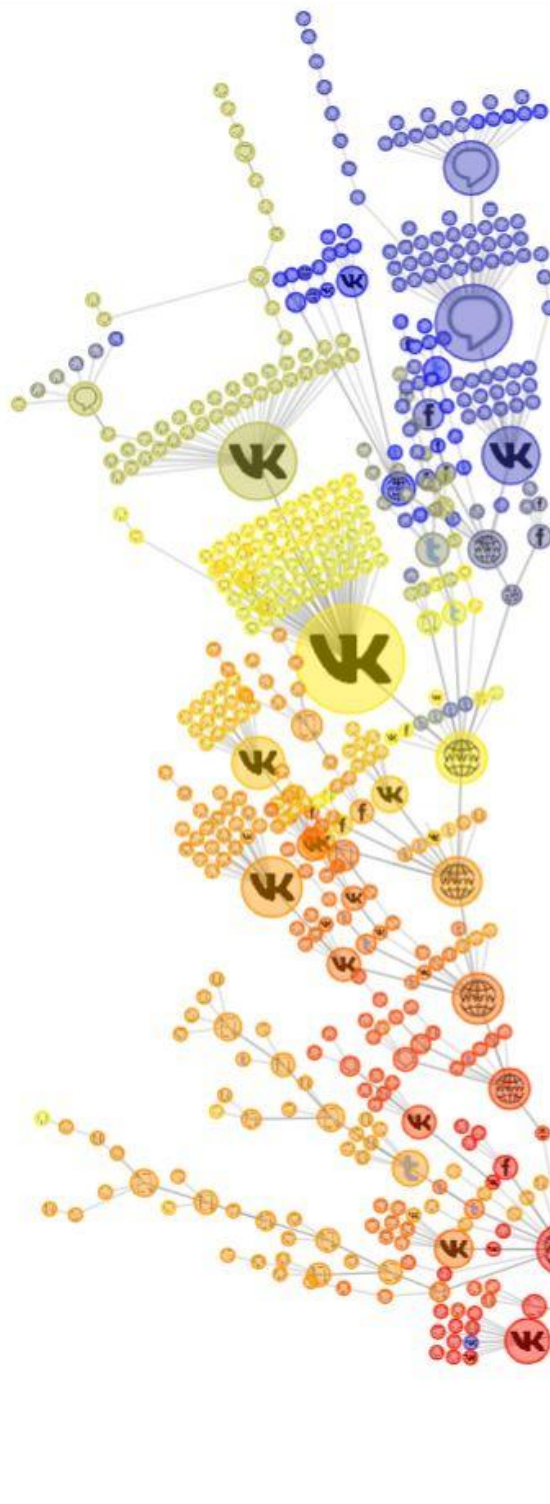
- анализ поведения аккаунта,
- автоматическое формирование портрета и графа реальных связей;
- анализ окружения аккаунта (по графу реальных связей);
- групповой анализ;
- автоматическое выявление наличие признаков риска в поведении аккаунта ;
- выявление аккаунтов по заранее описанному паттерну поведения.

Позволяют проводить анализ поведения аккаунтов пользователей социальных сетей, выявлять наличие признаков риска в поведении аккаунтов и сортировать полученные данные наиболее оптимальным образом.

«Крибрум. Сигнал»

Предназначена для определения источников и путей распространения информации в социальных сетях.

Позволяют анализировать характер распространения информации в социальных медиа, определять основные узлы распространения информации, естественный или искусственный





«Крибрум. Сигнал»

Реализует функции:

- поиска информации в хранилище данных Крибрум;
- определение первоисточников информации и цепочек ее распространения;
- визуализации характера распространения информации в социальных сетях.

СПАСИБО ЗА ВНИМАНИЕ!

АО «Крибрум»

Тел.: +7 499 372 5330

E-mail: info@kribrum.ru

Web: www.kribrum.ru



КРИБРУМ

Мы слушаем сеть