

Создание поисковых объектов в системе «Крибрум»

Школа прикладного анализа данных Data-Diving



Университетский
консорциум
исследователей
больших данных



КРИБРУМ
Мы слушаем сеть

Евгений Петров

Система «Крибрум»

«Крибрум» – система мониторинга и анализа социальных медиа.

Система собирает упоминания объекта (компания, персона, продукт компании) из русскоязычных ресурсов Интернета:

- социальные сети
- интернет СМИ
- блоги и форумы
- тематические порталы



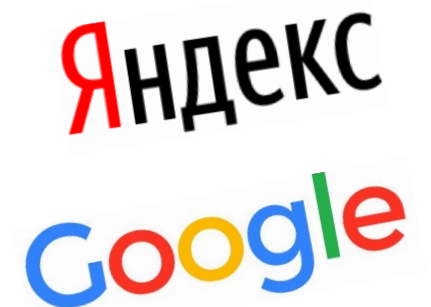
КРИБРУМ
Мы слушаем сеть

Поисковые объекты

- Поисковые объекты формируются на основе поисковых запросов, идентичных запросам в поисковых системах.
- Необходимо идти от полноты к качеству.

Например, в качестве объекта выступает **Юрий Алексеевич Гагарин**. Можно описать его как "**Юрий Алексеевич Гагарин**". Однако, при этом будут пропущены упоминания типа "**Гагарин**" (без имени и отчества), "**Гагарин Юрий**" (без отчества), "**Юрий Гагарин**" и т.д. Таким образом, необходимо сначала завести запрос "**Гагарин**", а потом по мере необходимости детализировать запрос.

"Гагарин"
"Юрий Гагарин"
"Гагарин Юрий"
"Гагарин Юрий Алексеевич"



Поисковые запросы

В запросах можно использовать:

- буквы алфавитов на основе кириллицы и латиницы.
- цифры
- пробел между словами
- операторы (кавычки, минус, логическое И, логическое ИЛИ, скобки)

Необходимо помнить:

- каждый запрос пишется в новой строке, все условия одного запроса пишутся в одной строке
- слова через дефис необходимо писать в кавычках, заменяя дефис на пробел. Например, «жар птица», «генерал майор» и т. д. При этом нельзя написать без кавычек (жар птица, генерал майор)
- если есть сомнения, что морфологические формы термина будут обработаны корректно, то нужно внести в поисковые запросы все словоформы термина. Например, «кэшбэк», «кэшбэку», «кэшбэка», «кэшбэком» и т. д.

Операторы поисковых запросов

Двойные кавычки:



Например "форд" - будут собраны сообщения содержащие данное слово в указанной форме

- Нельзя использовать прочие типы кавычек, например: «форд», “форд”.
- Для каждой кавычки обязательно должна быть парная!
- Внутри кавычек могут находиться только: буквы, цифры, пробелы между словами.
- Дефисов, точек и других знаков препинания внутри кавычек быть не должно – вместо них ставится пробел: "п в дуров"

Операторы поисковых запросов

Минус:

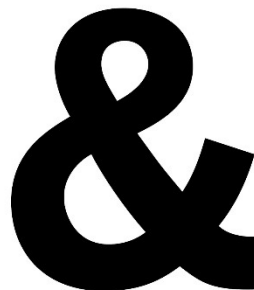


Например: **форд -харрисон** будут собраны сообщения содержащие слово «форд» в любой форме и не содержащие «харрисон»

- Минус можно использовать только перед словом для исключения сообщений, его содержащих.
- Категорически нельзя использовать его для поиска сообщений, содержащих дефис типа Санкт-Петербург или составные слова типа юго-западный. При необходимости найти подобные выражения нужно использовать кавычки и заменить знак "—" на пробел: " Санкт Петербург ".

Операторы поисковых запросов

Логическое И:



Например: **форд & фокус** будут собраны сообщения содержащие оба слова

1 – True, 0 - False

0 & 0 = 0 Продам новую Мазду 6.

1 & 0 = 0 Водитель на автомобиле **Форд** Мустанг не справился с управлением.

0 & 1 = 0 Сегодня проводится вечер магических **фокусов**.

1 & 1 = 1 Актуальные цены и стоимость различных комплектаций нового **Форд Фокус**.

Операторы поисковых запросов

Запрос: **форд фокус**

Будут собираться сообщения содержащие оба этих слова, однако, расстояние между ними не будет превышать 3х слов.

Например, **соберется:**

- Форд показывает фокусы;
- Автомобиль форд фокус.

Не соберется:

- Компания Форд представила новую модель фокуса.

Расстояние можно указать с помощью слеша «/».

Например: **(форд фокус) /2**.

Операторы поисковых запросов

Логическое ИЛИ:



Например: **фокус | focus** будут собраны сообщения содержащие хотя бы одно из слов

1 – True, 0 - False

0 | 0 = 0 Компания «Ford» объявила о выпуске нового автомобиля.

1 | 0 = 1 Как настроить **фокус** при фотосъемке?

0 | 1 = 1 Новая модель Ford **Focus** 2019 года.

1 | 1 = 1 Тест-драйв «**Focus** без **фокусов**».

Скобки **()**

Например: **(фокус | focus) & ford**

Интерактив

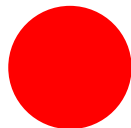


Университетский
консорциум
исследователей
больших данных

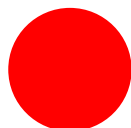


Национальный
исследовательский
Томский
государственный
университет

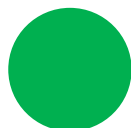
«ТОМСК»



“ТОМСК»

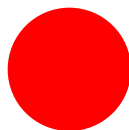


"ТОМСК"

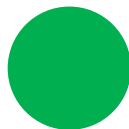


Интерактив

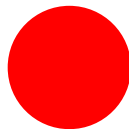
"Ростов-на-Дону"



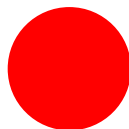
"Ростов на Дону"



`Ростов на Дону`



«Ростов на Дону»



Какое сообщение соответствует запросу?



(Наполеон | "император Наполеон") –пирожное –торт

- Наполеон - слоёный торт или пирожное с кремом.
- Наполеон – французский полководец и император.
- Император тортов - торт Наполеон

Какое сообщение соответствует запросу?



Университетский
консорциум
исследователей
больших данных



Национальный
исследовательский
Томский
государственный
университет

(Наполеон & торт)

- Наполеон - слоёный торт или пирожное с кремом.
- Наполеон – французский полководец и император.
- Император тортов – торт Наполеон

Какое сообщение соответствует запросу?



(тгу & вуз)-тамбов-тамбовский

- Сейчас в ТГУ идёт стратегическая сессия, где представители администрации, вузов и НИИ обсуждают возможности кооперации.
- В стенах вуза ТГУ (Тамбовского государственного университета имени Г.Р. Державина) насыщаются знаниями около 10 тысяч студентов.
- Высшая школа журналистики ТГУ проводит конкурс студенческих работ

ě j ə ě z' əj̃ каGЗ ʃ' - ê аG

Турнир по киберспорту

("мягков"|"гойко"|"мягкова"|"мягковым"|"мягкове") &
турнир
"турнир" & "университетская лига"
"национальная ассоциация университетского киберспорта"
"национальной ассоциации университетского киберспорта"

Томский государственный университет

"государственный университет"&(томск|томский)
"tomsk state university"
"tomskuniversity"
"tomskstateuniversity"
"томским императорским университетом"
"томским государственным универом"
"tsu ru"
"tsuru" -"nissan"
"тгу" -тольятти -тамбов -тольяттинский -тамбовский -тюмень -тула -"державина"

Экстремистские сообщества

"джихад"
"муджахид"|"муджахед"|"моджахед"
"рибат"
"джахилия"|"джахилийя"|"джахилийский"
"чеченский джамаат"|"джамаат галгайче"

d'oĭĕ ηĕ; ^ j ĭĕə3



Национальный
исследовательский
Томский
государственный
университет

Отчеты Сообщения 787 Авторы 489 Задания Проект: Томский Государственный Университет

Всё Избранное / Остальные Прочитанные / Непрочитанные С заметками С заданиями С ответами Пересечение Удаленные

↓ по дате публикации Экспорт Тексты Дубли Подсвечивать теги Только новые + Добавить 1-250 из 701

TV TOMSK TVTOMSK.RU/ВЕСТИ-ТОМСК 14:05:00, 12 ноября

В ТГУ в эти дни проходит международный конкурс...

В ТГУ в эти дни проходит международный конкурс пианистов. Это четвертый Сибирский международный конкурс пианистов имени Фридерика Шопена. Обладатель гран-при конкурса сможет выступить с двумя концертами в Польше, на родине великого пианиста. Международный конкурс пианистов проходит в ТГУ www.tvtomsk.ru

В ТГУ в эти дни проходит международный конкурс пианистов. Это четвертый Сибирский международный конкурс пианистов имени...

https://vk.com/wall-39787587_62182 1

Объект: Томский Государственный Университет

Категории: Категория

Теги: Tag

Европа / Польша, Европа / Россия / Сибирь

Обнаружено 12 ноября 2019, 14:20:17

Отправить Дать задание Добавить заметку

Ахерон--Страшные истории 14:00:00, 12 ноября

27-летний житель Томска получил крайнее...

... российских ученых. Этот аппарат был создан в Томском

Показать полностью... государственном университете (ТГУ) и о нем еще в 2015 году сообщали российские СМИ. Однако лишь недавно его решили испытать на человеке и более того, на реальном пострадавшем

Как сообщает ТАСС, аппарат лечит обмороженные части тела с помощью слабого СВЧ-поля, очень быстро прогревая пострадавшую конечность. По словам специалистов ТГУ, аналогов этому аппарату в мире нет, это уникальное российское изобретение. После того как ...

© 2010-2019 Крибрум Правила использования

Параметры по умолчанию

Ключевые слова

Поиск в найденном

Период сегодня неделя месяц

6 - 12 ноября 2019

время публикации

время обнаружения

Объекты: выбрать

Томский Государственный Университет

Источники: выбрать

Социальные сети

Отчёт по проекту "Томский Государственный Университет"			
7			
8			
9	Период:	1 Ноябрь 2019 - 7 Ноябрь 2019	
10	Объекты:	Томский Государственный Университет	
11	Источники:	vk.com	
12	Типы источников:	Только	
13	Роли:	Все	
14	Оценки:	Все	
15	География:	Все	
16	Категории:	Все	
17	Теги:	Все	
18	Авторы:	Все	
19			
20	За отчетный период:		
21	Количество сообщений:	799	Детальный отчет на листе "Сообщения"
22	Изменение к предыдущему периоду:	-172 (-17%)	
23	Максимум сообщений в сутки:	165	5 Ноябрь 2019
24	Количество авторов:	413	Детальный отчет на листе "Авторы"
25	Активность (постов на автора):	1,93	
26	Количество источников:	1	Детальный отчет на листе "Источники"
27	Тональность упоминаний:	798 (99%) нейтральные	Детальный отчет на листе "Тональность"
28		0 (00%) позитивные	
29		1 (00%) негативные	

Демонстрация:

<https://kribrum.io/>

- Познакомились с системой «Крибрум»;
- Изучили такие понятия, как поисковый объект и поисковый запрос;
- Научились формировать поисковые запросы;
- Изучили операторы, которые используются при формировании запросов;
- Рассмотрели примеры поисковых объектов;
- Познакомились с интерфейсом системы.

Спасибо за внимание

Евгений Петров, petrov@data.tsu.ru



Университетский
консорциум
исследователей
больших данных



КРИБРУМ
Мы слушаем сеть