

УЧЕБНЫЙ ЦЕНТР «ТОПЭКСПЕРТ»



ТОПЭКСПЕРТ

УЧЕБНЫЙ ЦЕНТР

Модуль кластеризации запросов

10 поток профессионального курса по SEO

Работу выполнил: Юрий Игнатов

Дипломный руководитель: Дмитрий Иванов

Учесть: Пороги совпадений разные для каждого порога!

- Описание проекта

Модуль кластеризации запросов на основе топов.

Разбивает запросы на кластеры, по мнению топов Яндекса.

- Задачи, которые решает модуль

- Сегментация семантического ядра
- Расширение семантического ядра
- Определение разделов, категорий сайта

- Входные данные (описание и таблица)

- Запрос
- Регион

- Выходные данные (описание и таблица)

Файл, состоящий из одного столба кластеров.

Пример:

Дайвинг клуб
Дайвинг центр

Дайвинг оборудование
Снаряжение для дайвинга

...

- Формулы

Переменные:

«Запрос1-url1» – первый запрос с адресом сайта, далее по таблице БД.

«Кластер1» - первая группа запросов.

«N» - следующие значения.

Истина:

Запрос1-url1 смотрим в КЭШе все ЗапросN-urlN ≥ 4

кластер1 = все совпавшие запросы;

Лож:

Запрос1-url1 ≤ 3

Кластер2 = Запрос1

Автор: Юрий Игнатов, 10-ый поток курсов ТопЭксперт

Дипломный руководитель: Дмитрий Иванов

Далее берем следующий «Запрос-url2» ищем совпавшие варианты (запрос и url), по выше описанному алгоритму.

- Список модулей, с которыми взаимодействует модуль

?

- Описание процессов взаимодействия

1. **Парсинг** – по каждому запросу собираем URL`ы топ 10;
2. **Кластеризация** – каждый «запрос-url» сравниваем со всеми «запросN-urlN». Если совпадений ≥ 4 , добавляем «запрос» в проверяемый кластер. Если совпадений ≤ 3 , проверяемый «запрос» добавляем в отдельный кластер. В конце списка запросов каждого кластера дописываем двойной перевод строк.
3. **Файл** – Данные сохраняем в текстовый файл, содержащий один столб кластеров.

Осуществление каждого пункта процессов выполняется автоматически, в случае сбоя в ручном режиме можно продолжить процесс без потери данных!

- Карта логических связей выполнения модуля

1. Ввод исходных данных;
2. Обращение к Яндексу;
3. Сбор данных, сохранение в БД;
4. Обработка, сохранение в БД;
5. Сохранение в файл.

- Предполагаемая нагрузка

?

- Особые требования

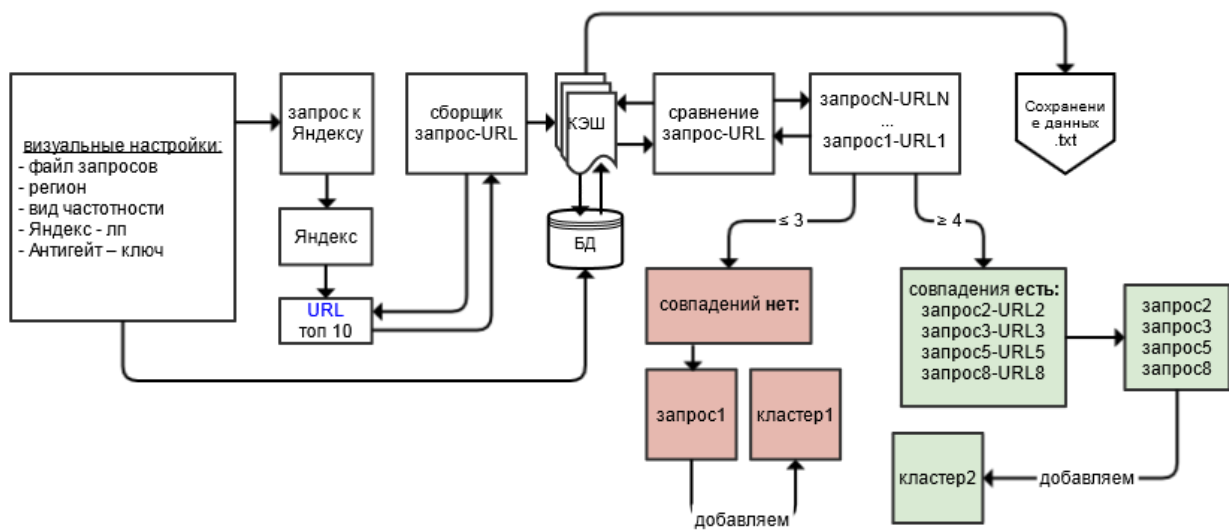
При многократном обращении, Яндекс выдает капчу

Решение: подключить антигейт, предложить еще варианты!

- Процесс остановки модуля

1. Окончание обработки данных;
 2. На стороне Яндекса;
 3. На стороне сервера;
- Вывести окно с сообщением вариантов устранения проблем.

- Процесс запуска модуля



- Формирование бекапов

Данные проектов хранятся в базе данных:

- Данные собранные с выдачи Яндекса (Сбор данных).
- Кластеризированные данные проекта (Кластеризация).

В случае сбоев, сбор данных или кластеризацию можно восстановить.

- Восстановление бекапов

- Через хостинг провайдера.
- После: Сбор данных / кластеризация / файл
Предлагается сохранить проект в .TXT

- Предполагаемое расширение модуля

?

- Возможные причины поломки модуля

?

- Работа модуля в случае поломки на каждом участке

?

Компонент: Обзор визуального меню (закладок)

1. Закладка «Сбор данных»:

- 1.1. В начале работ проекту необходимо дать название или выбрать из списка проектов.
- 1.2. Выбираем TXT файл с запросами (все запросы в столб)
Скрипт очищает пропуски строк;
- 1.3. Выбираем частотность: **общая**, **""**, **“!”**;
- 1.4. Запускаем.

Если все данные не заполнены, выдает ошибку с указанием не заполненного поля.

2. Закладка «настройки»:

- 2.1. Выбираем регион;
- 2.2. Яндекс лк - логин:пароль (поле для ввода);
- 2.3. Антигейт – ключ (поле для ввода);

3. Закладка «Бэкап»:

В виде списка из 4-х столбов:

название проекта	сбор данных	кластеризировать	Скачать	БД
dreamdive.ru	Собрать	собрать	файл	удалить
	Продолжить	продолжить		
	Обновить	обновить		

собрать	Если данных нет выводим
продолжить	<ul style="list-style-type: none"> • Если произошла ошибка, процесс останавливаем и выводим данную ссылку. • При нажатии на ссылку сбор или кластеризация продолжается с запроса, на котором остановлен процесс. • Если проблема не решена, выводится сообщение о вариантах решения.
обновить	<ul style="list-style-type: none"> • После успешного сбора или кластеризации выводим ссылку. • Если необходимо обновить данные, после нажатия происходит повторный сбор данных или кластеризация.
файл	Скачать кластеризированные данные в файл.
удалить	Из базы данных по проекту.