“**Разработка рекомендательной системы по размещению наружной рекламы”**

**Разработал и подготовил Лученков И.В, +79119319719,** [**il.v.luchenkov@gmail.com**](mailto:il.v.luchenkov@gmail.com)

**Описание.**

Разрабатываемая платформа осуществляет интерактивное отображение на географической карте (слои OpenStreeMap, формат leaflet) города Санкт-Петербург и ЛО вектора(направления) наиболее перспективного размещения наружной рекламы (билборд), относительно потребительский предпочтений мобильных абонентов при помощи диаграммы Венна.

**Цель.**

~~Выдавать рекомендательный аналитический отчет заказчику (рекламное агентство) по наиболее перспективному содержанию и размещению наружной рекламы на билбордах для таргетированного продвижения услуг в режиме реального времени.~~

Платформа позволит оптимизировать планирование и управление рекламными кампаниями, автоматизирует их запуск, обеспечит онлайн-мониторинг эффективности размещения рекламы.

**Ограничения.**

Доступ к raw данным сотового оператора.

Не проработана конверсия по абонентский кликам.

Пилот развернут на Centos7.3, 16 RAM, RStudioServer(Docker).

**Архитектура.**



**Задачи.**

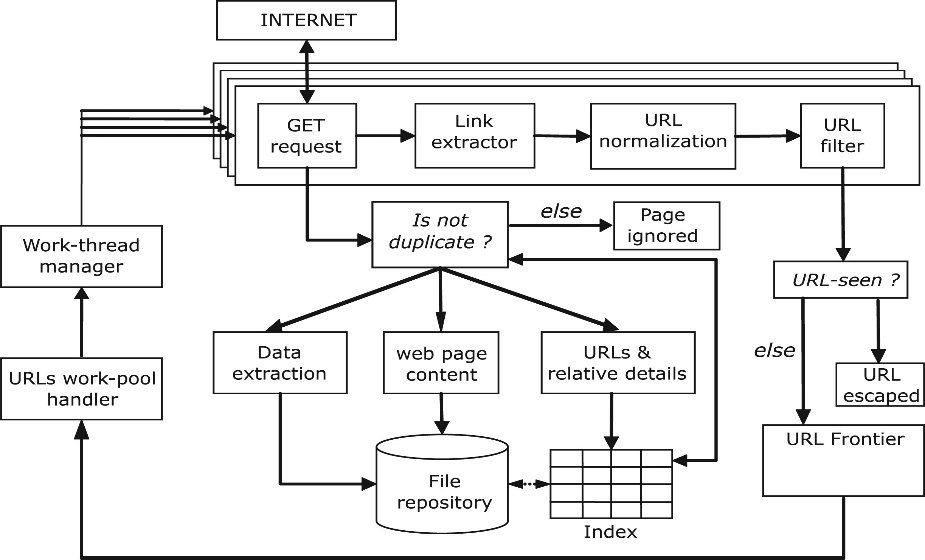
**Фаза 1. Консолидация файлов**

1. Консолидация данных, узел PCEF. Данные содержат “http” запрос абонента;
2. Консолидация данных, узел SAE. Данные содержат “msisdn” и “IP” абонента;
3. Консолидация данных, узел MME. Данные содержат “msisdn” абонента и “sector” BS;
4. Формирование файла [cell.txt]. Файл содержит “sector” BS и полицейский адрес.

**Таблица 1. Метрики потоков**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **n** | **Узел** | **Объем файла, мБ** | **Ср.Кол, 5 мин** | **Объем мБ за 5 мин.** |
|  |  |  |  |  |
| **1** | PCEF x 8 | 50 | 8 | 2.000 |
| **2** | SAE | 41 | 30 | 1.230 |
| **3** | MME | 10 | 115 | 1.150 |
| **4** | Cell.txt | 7.3 | const | const |

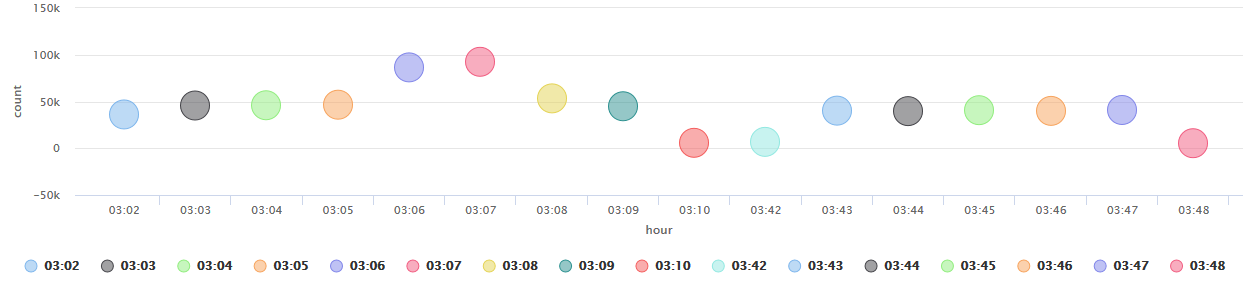
**Фаза 2. Краулинг (crawling) и скрапинг (scraping), PCEF.**

****

**Рис.2 Алгоритм**

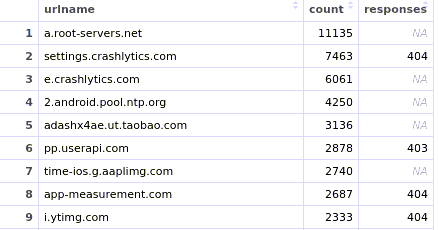
1. Фильтр (drop) по hostname, содержащих название:

"cloud|mail|apple|google|yandex|facebook|vk.com|avito|instagram|ya.ru|ssl.|viber|microsoft|youtube|skype|radio|amazon|localhost……

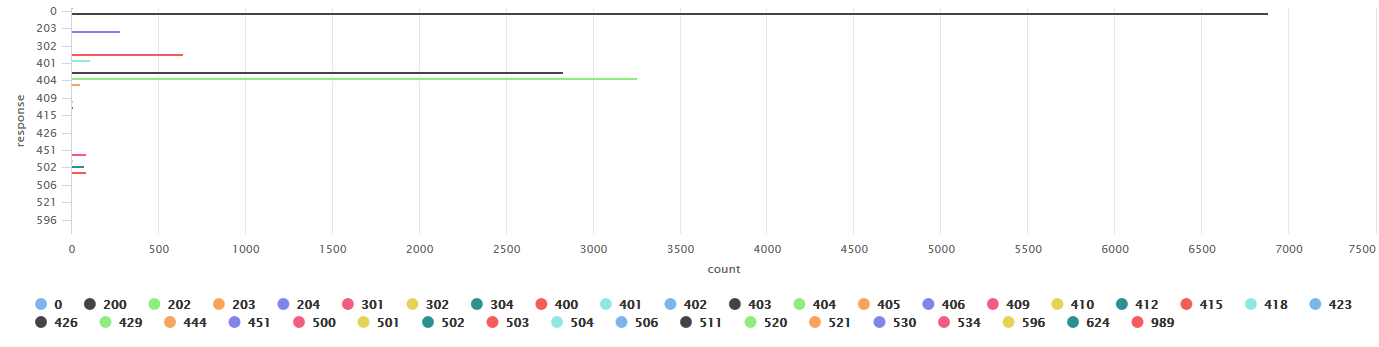


**Рис.3 Абонентская активность после применения фильтра за период [03:02 – 03:48]’P.M.**

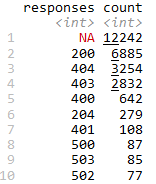
1. Анализ url для формирования вектора с кодами ответов. Полученные ответы сохраняются в бинарный файл [responses.RData]. При наличии в новых данных информации о новых url производится опрос указанного url и автоматическое дописывания бинарного файла новыми url с полученным кодом ответа.



**Рис.4 R’dataFrame top\_url, вычитанный из бинарного фала [responses.RData].**

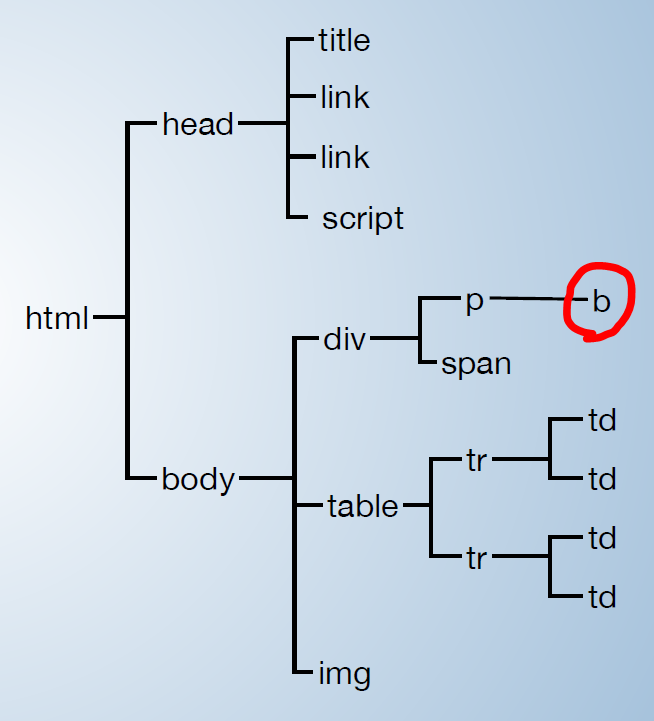


**Рис.5 Распределение URL кодов.**



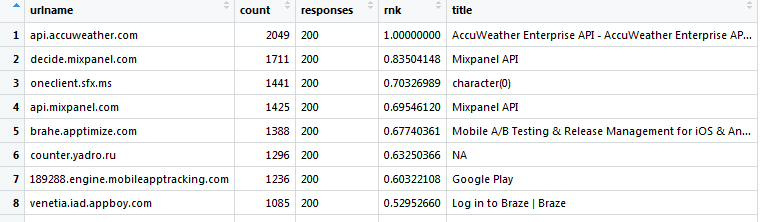
**Рис.6 Top 10 URL кодов.**

1. Фильтр (keep) URL по коду 200;
2. Page Rank, min = 0, max = 1;
3. Извлечение титула и получение информации по клику абонента (в HTML5–это узел “b”);

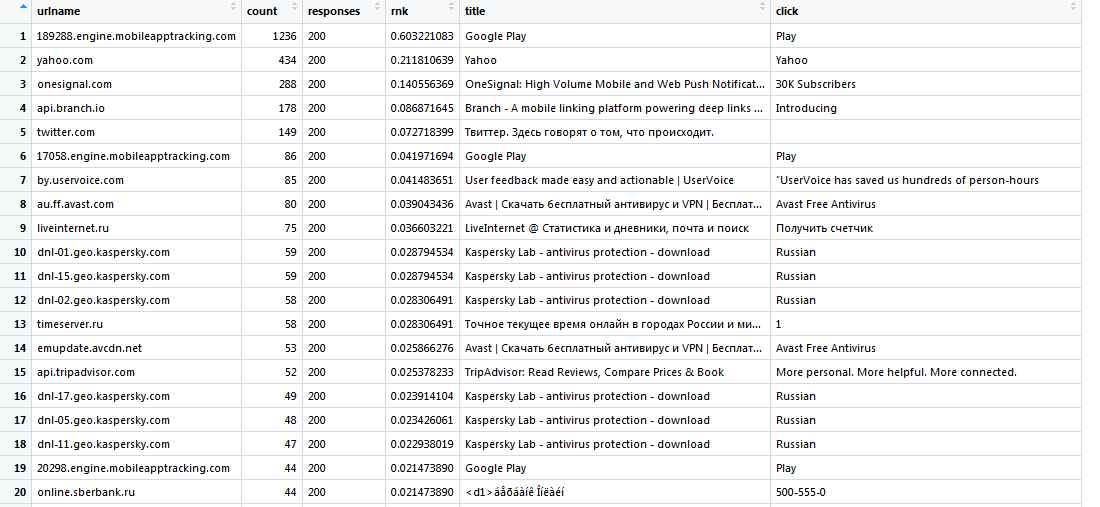


1. По результатам п.5 формируются 2 бинарных файлы:

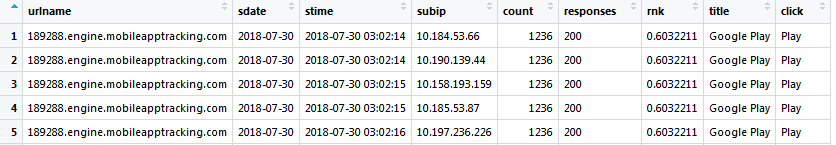
[titles.RData](~8000 url time in sec elapsed 5351.009)

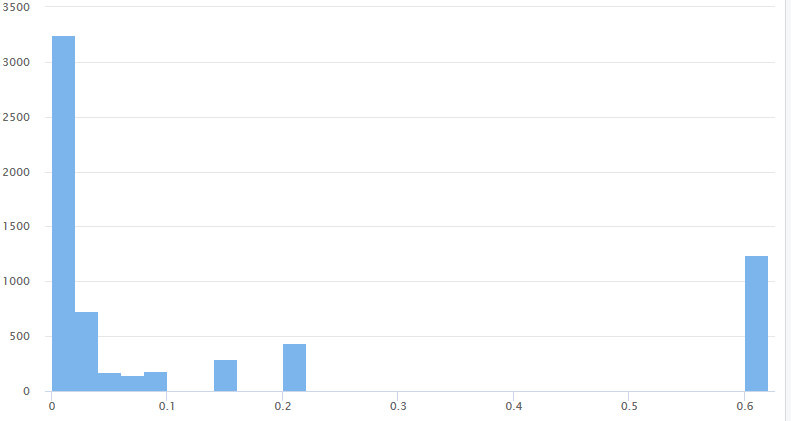
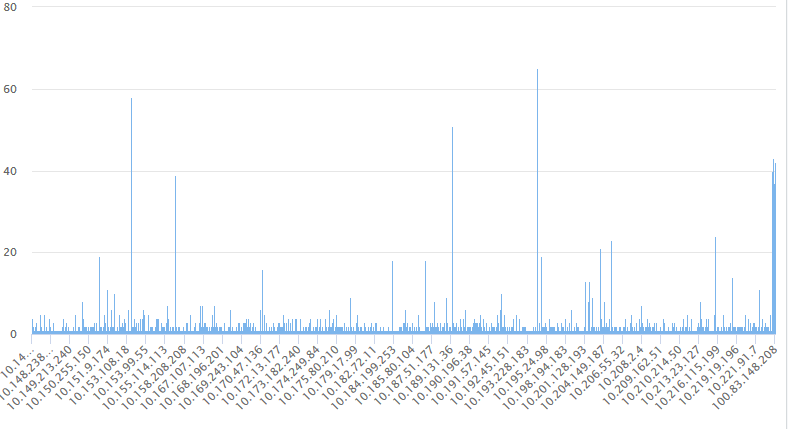


[click.RData] (~5000 url time in sec elapsed 3351.259)



1. Текст майнинг ?? (Не проработана конверсия по абонентский кликам.)
2. Формирование результата

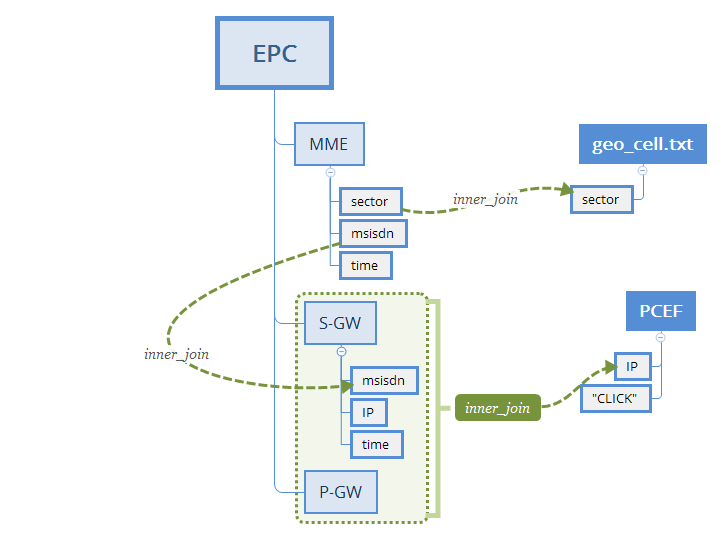




**Рис.7 Слева: Активность абонетов, справа: распределение по рангу.**

**Фаза 3. Объединение информационных элементов(ИЭ).**

1. Объединение ИЭ по первичным ключам: sector;msisdn;IP.



**Фаза 4. Отображение на карте.**

Концентрация базовых станций с фильтром по PageRank

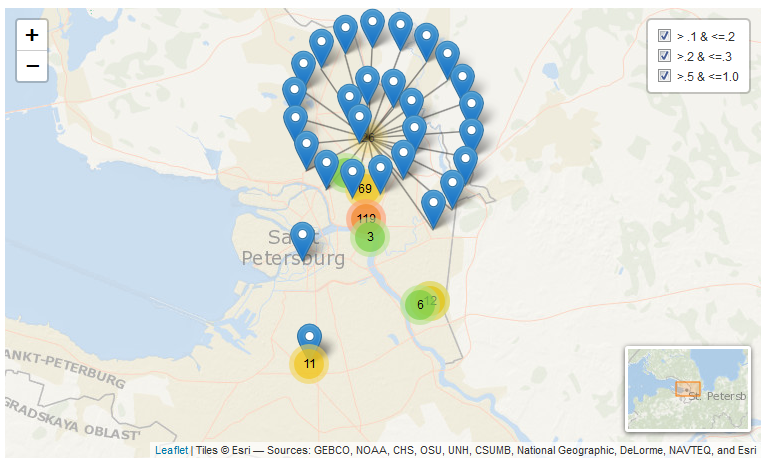


Диаграмма Венна. Определяем направление наиболее выгодного размещения наружной рекламы относительно потребительских предпочтений.

