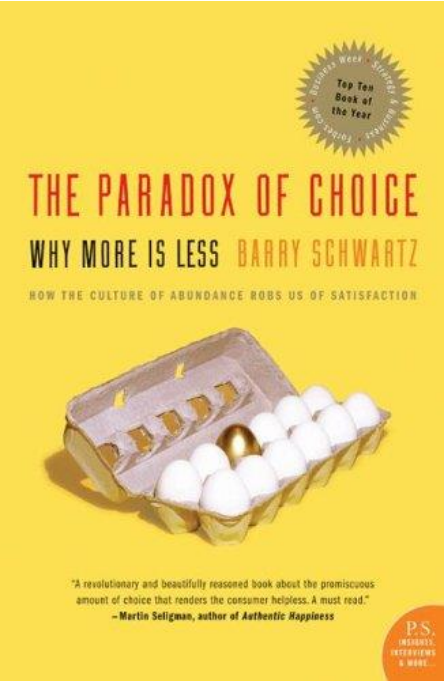
Технологи рекомендаций и нативной рекламы MyWidget

Алексей Воропаев, руководитель группы рекомендательных систем

Информации становится слишком много



«Парадоксальным образом изобилие ассортимента товаров и услуг ограничивает свободу нашего выбора.»

Многие уже научились на этом зарабатывать

NETFLIX 75% от 6.78MM $

amazon 35% от 107MM $

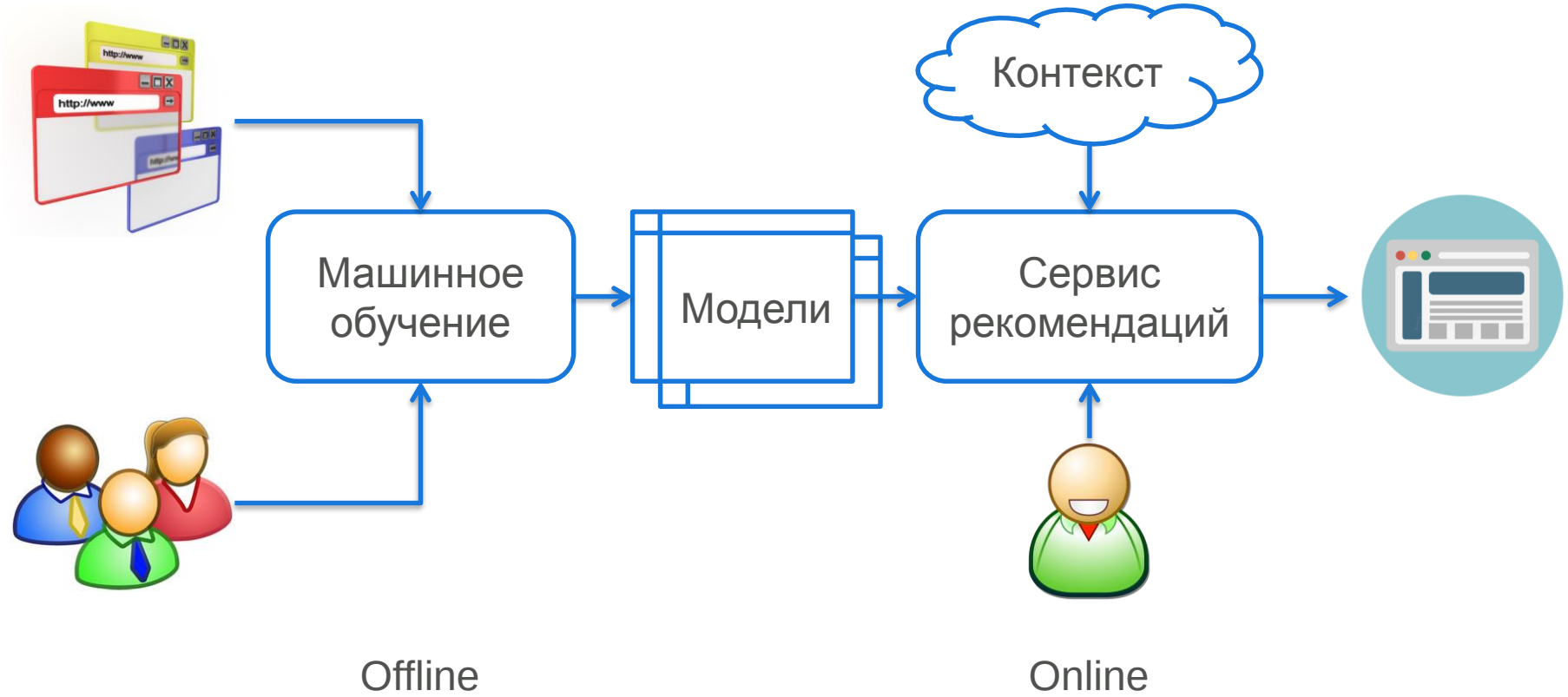
Немного про данные

5 TB/сутки

500 GB/сутки

100 GB/сутки

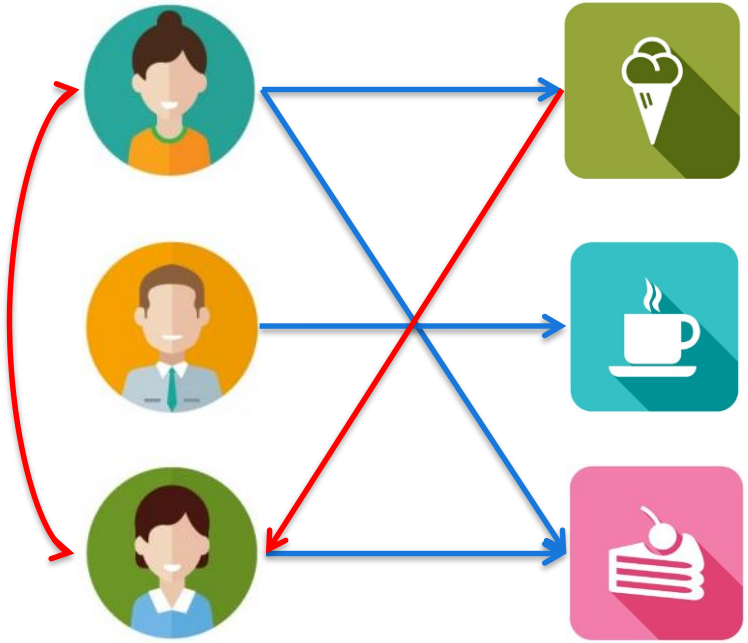
Схема работы сервиса



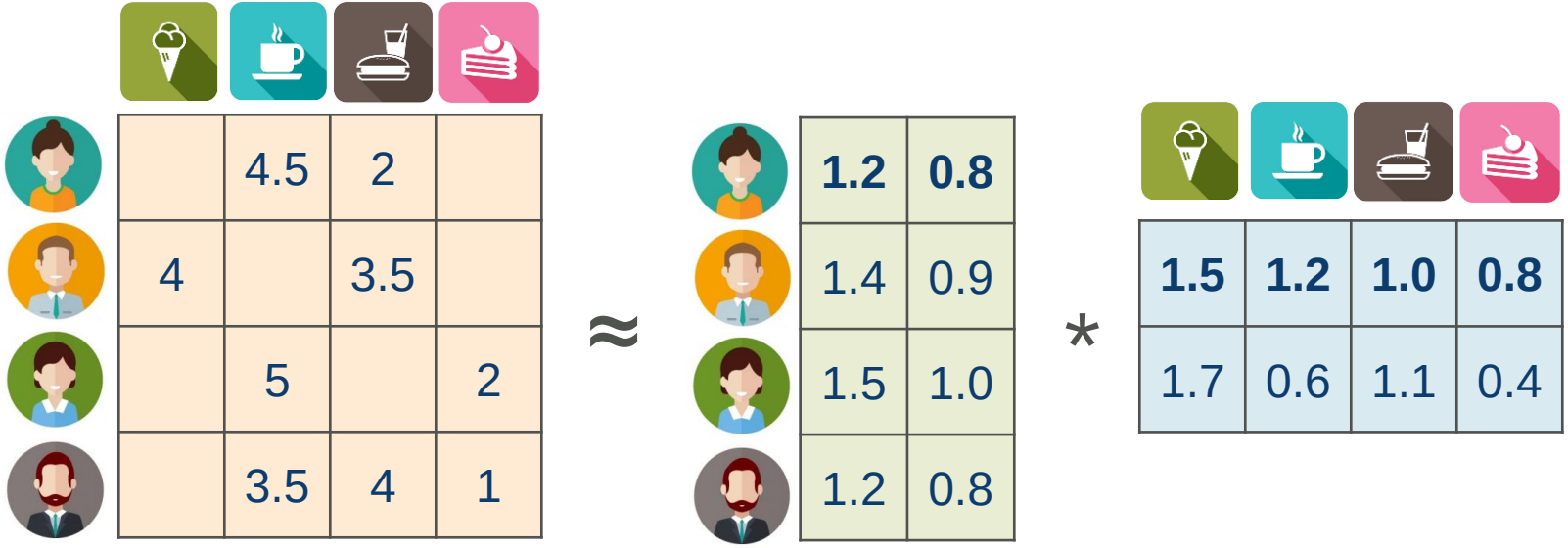
Основные подходы к рекомендациям

- Коллабаративная фильтрация
- Контентные рекомендации
- На основе свойств объектов
- На основе сообществ людей
- Гибридные системы

Коллабаративная фильтрация

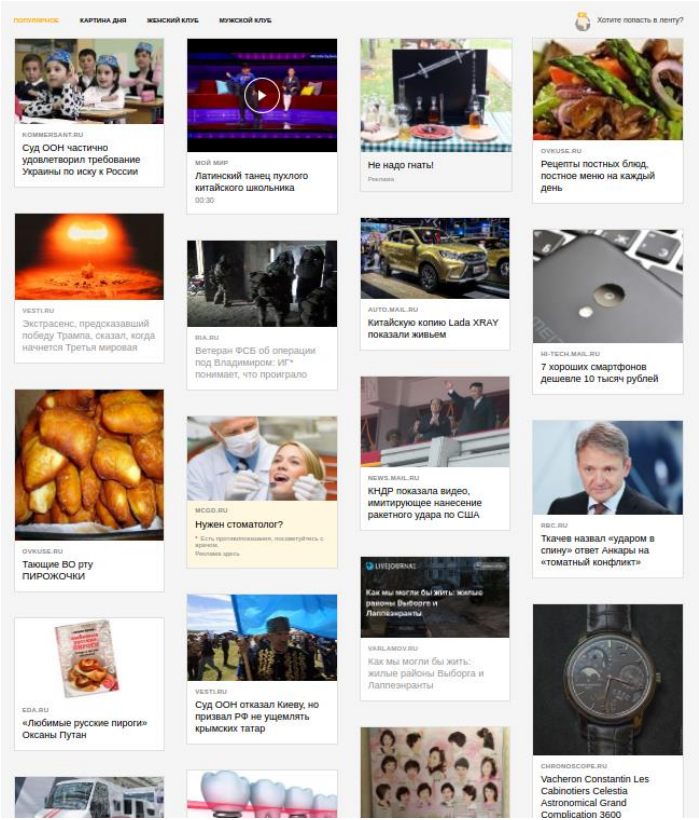


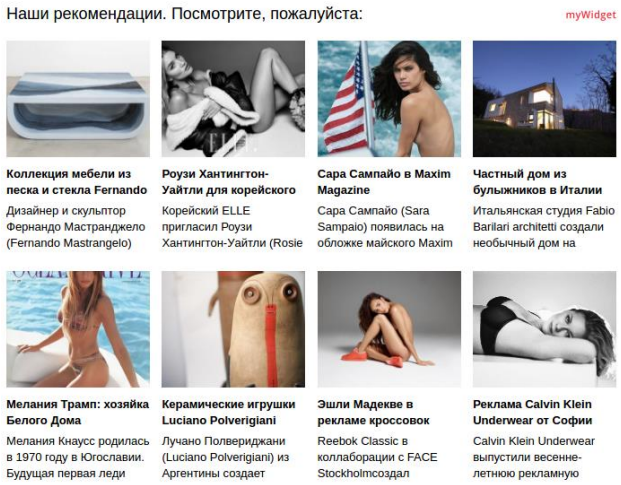
Матричное разложение



$$f(U, M) = \sum_{(i,j)\in I} (r_{ij} - \mathbf{u}_i^T \mathbf{m}_j)^2$$

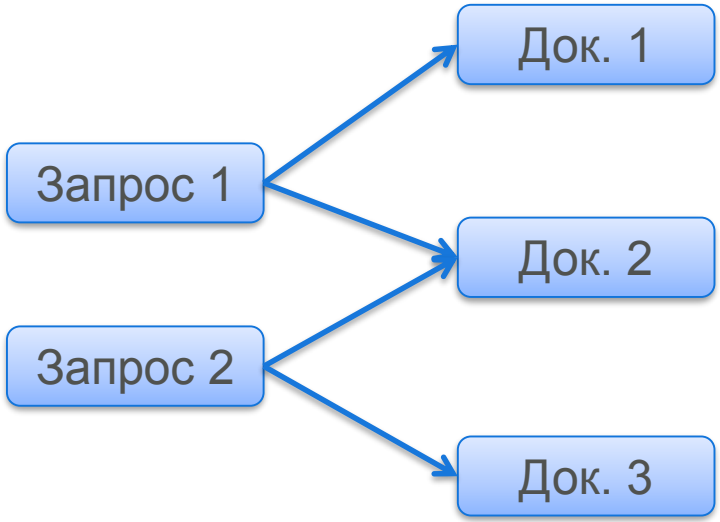
Персональные рекомендации

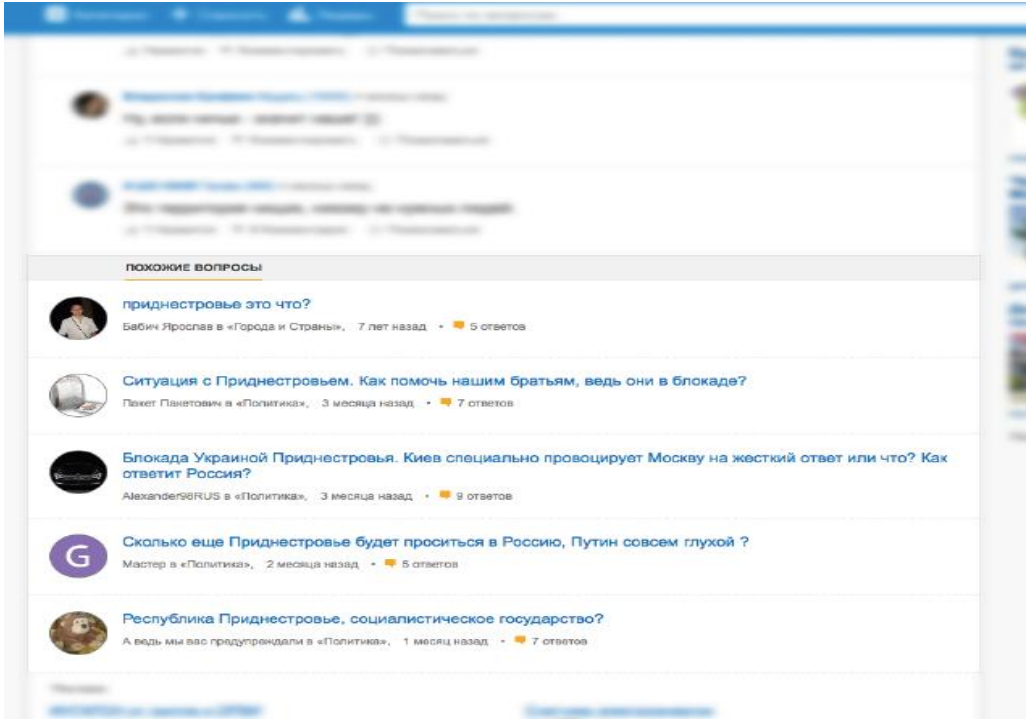




http://go.mail.ru/

Граф запрос-документ



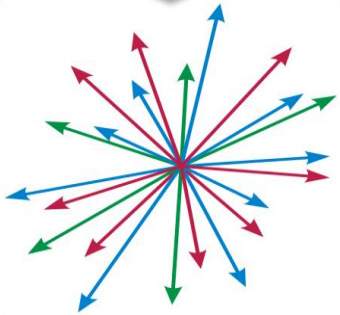


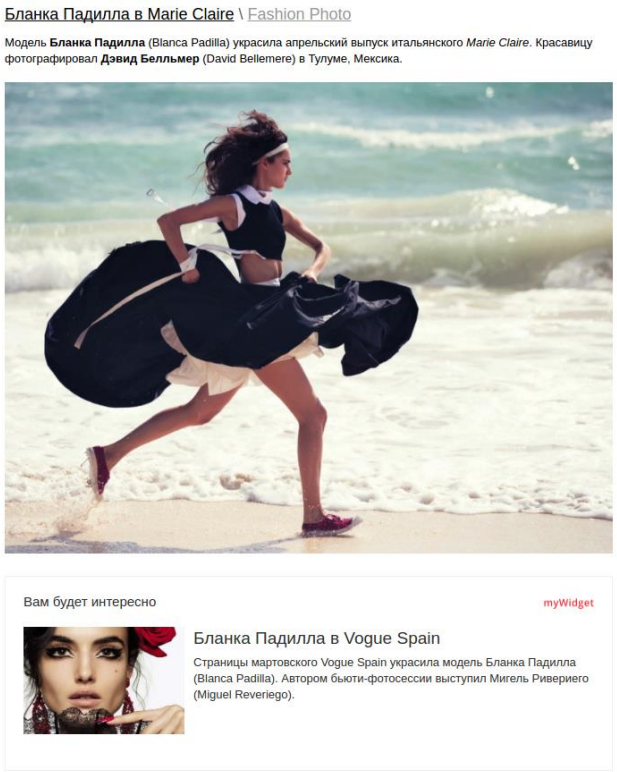
Контентные рекомендации

docs

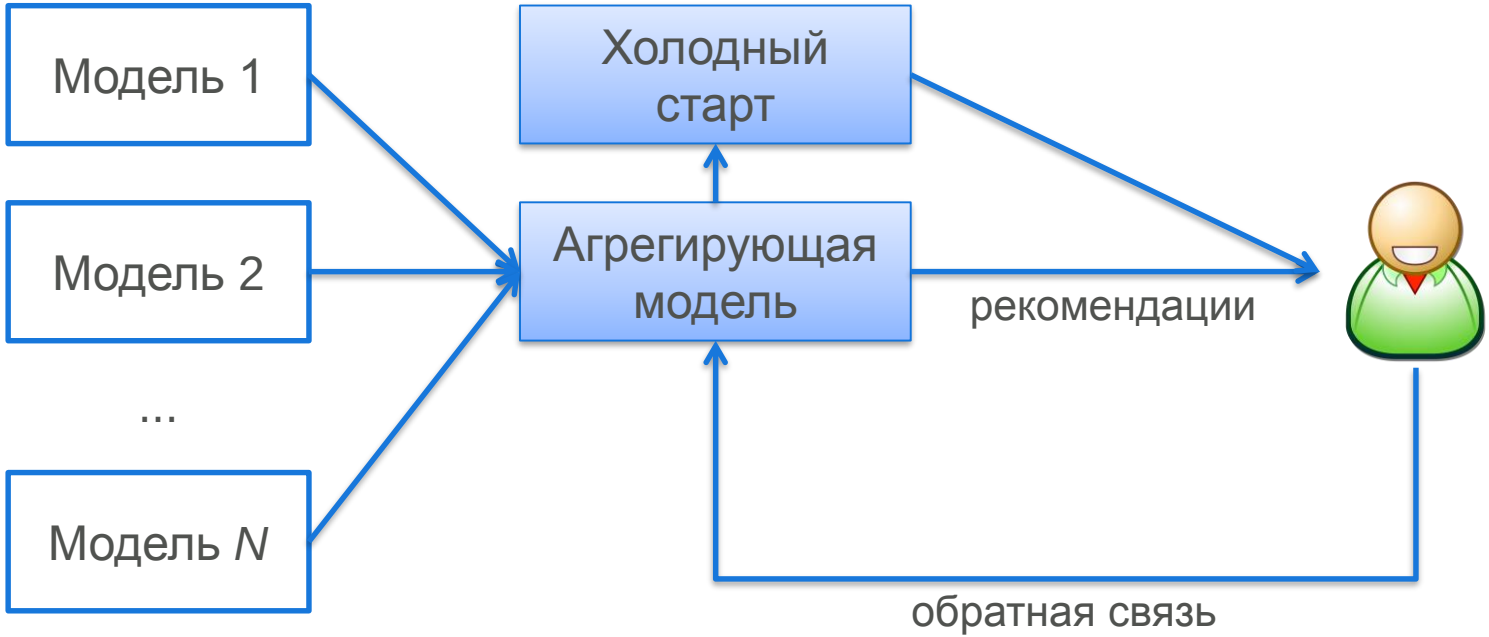
keywords

word2vec



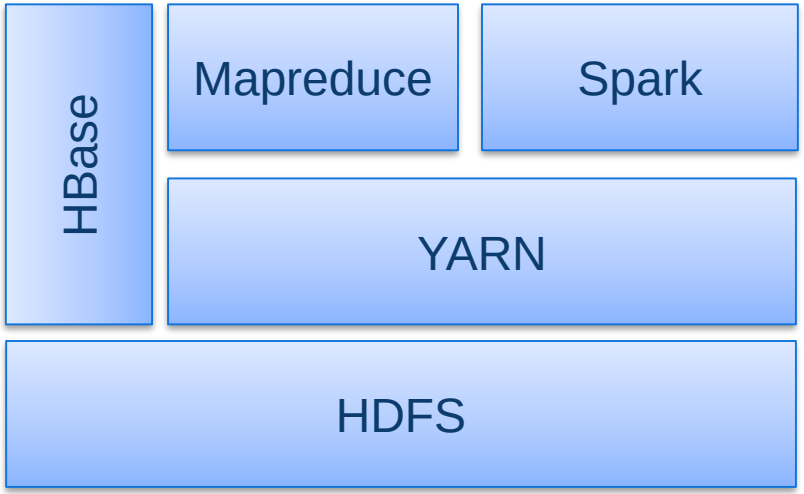


Гибридные модели

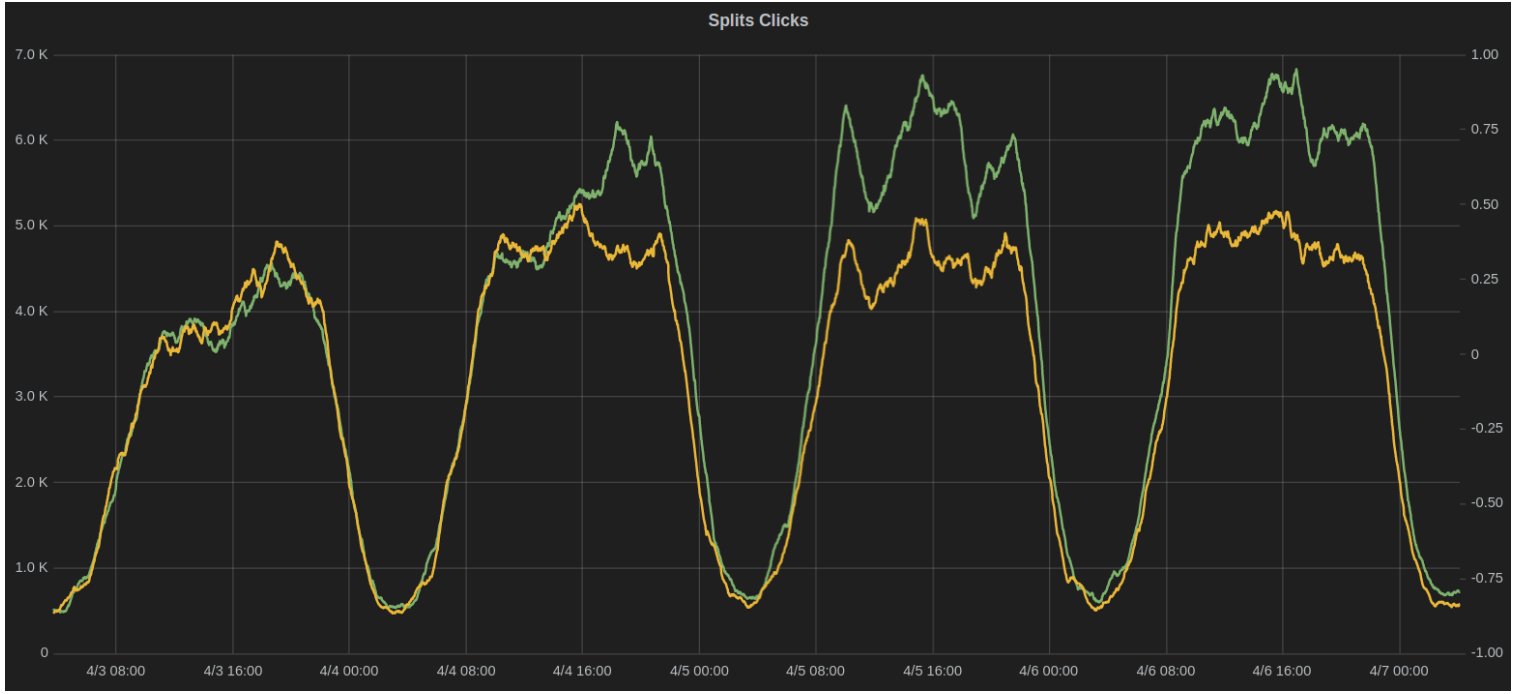


Hadoop

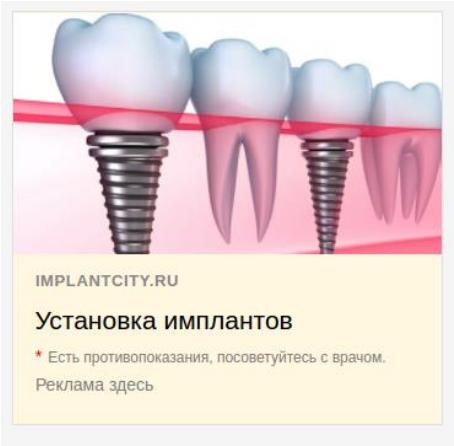
18 PB HDD
40 TB RAM

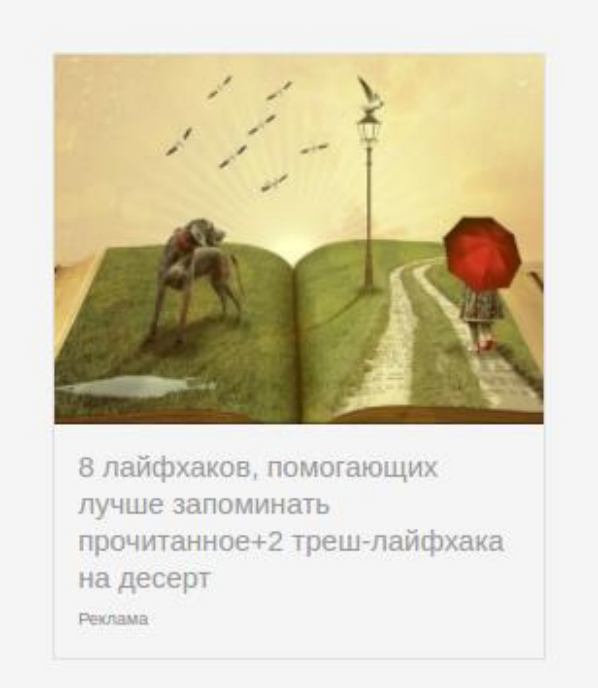


A/B тесты


Splits Clicks
7.0 K
6.0 K
5.0 K
4.0 K
3.0 K
2.0 K
1.0 K
0
1.00
0.75
0.50
0.25
0
-0.25
-0.50
-0.75
-1.00
4/3 08:00
4/3 16:00
4/4 00:00
4/4 08:00
4/4 16:00
4/5 00:00
4/5 08:00
4/5 16:00
4/6 00:00
4/6 08:00
4/6 16:00
4/7 00:00

«Нативная» реклама

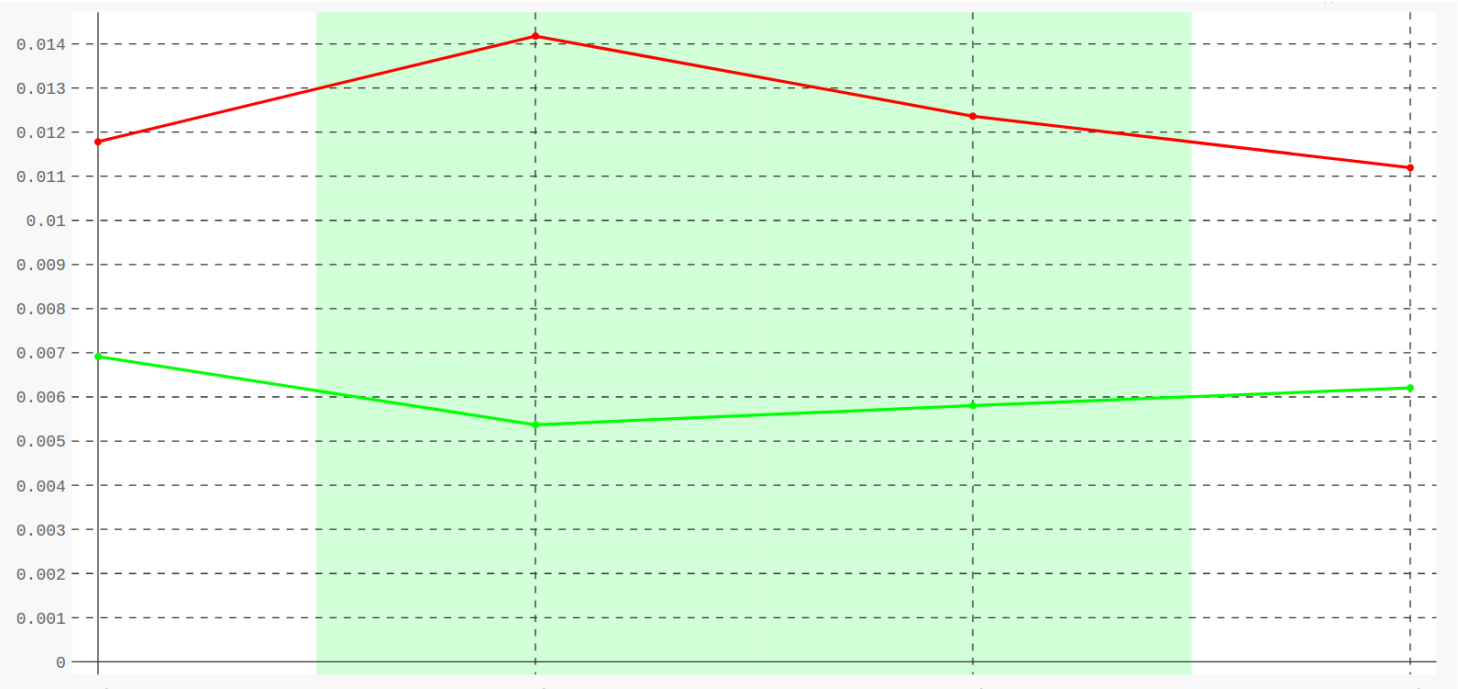




- Меньше раздражения
- Больше CTR

- Но, сложнее создавать

«Нативная» реклама



С точки зрения рексистемы

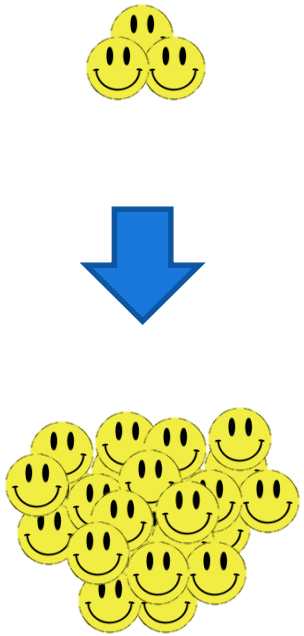
- Управления площадками размещения
- Побдор банеров для пользователя
- Оценки CTR для аукциона

- Кабинет рекламодателя
- Управление ограничениями показа
- Аукционная система (аукцион второй цены)

В чем разница

1. Нужно минимизировать прогрев
2. Нужно понимать, когда рекламу показывать не нужно

Look-alike



- Вектор сайтов
- Латентные факторы
- Контентные факторы
- Соц.дем факторы

Мы вам еще покажем!
widget.my.com

voropaev@corp.mail.ru