

Многокритериальная оптимизация поисковой выдачи в Авито

Андрей Дроздов



HighLoad⁺⁺

Профессиональная конференция
разработчиков высоконагруженных
систем

Проблема интересов



Примеры

- **запрос “ягуар”** — машина, животное или алкоголь=)
- **платные услуги** — байеры хотят органическую выдачу, монетизация хочет платную
- **новые услуги** — запчасти с доставкой в регионы

Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Пианино Korg

Пианино Yamaha

Детский синтезатор

Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Сегодня. Rel=1.0



Пианино Korg

Пианино Yamaha

Детский синтезатор

Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Сегодня. Rel=1.0

Пианино Korg

Вчера. Rel=0.9

Пианино Yamaha

Детский синтезатор

Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Сегодня. Rel=1.0

Пианино Korg

Вчера. Rel=0.9

Пианино Yamaha

Сегодня. Rel=0.8

Детский синтезатор

Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Сегодня. Rel=1.0

Пианино Korg

Вчера. Rel=0.9

Пианино Yamaha

Сегодня. Rel=0.8

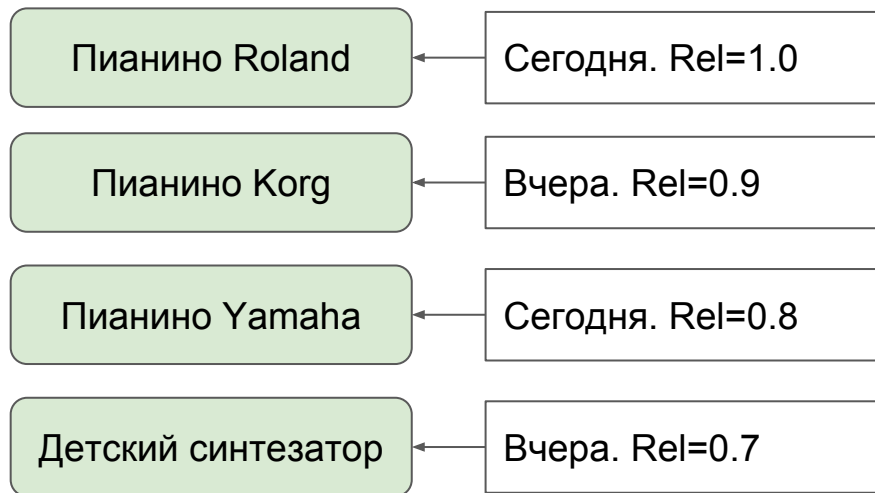
Детский синтезатор

Вчера. Rel=0.7

Общая проблема

Любая категория ▾	пианино	По всей России ▾	Найти
-------------------	---------	------------------	-------

Buyer



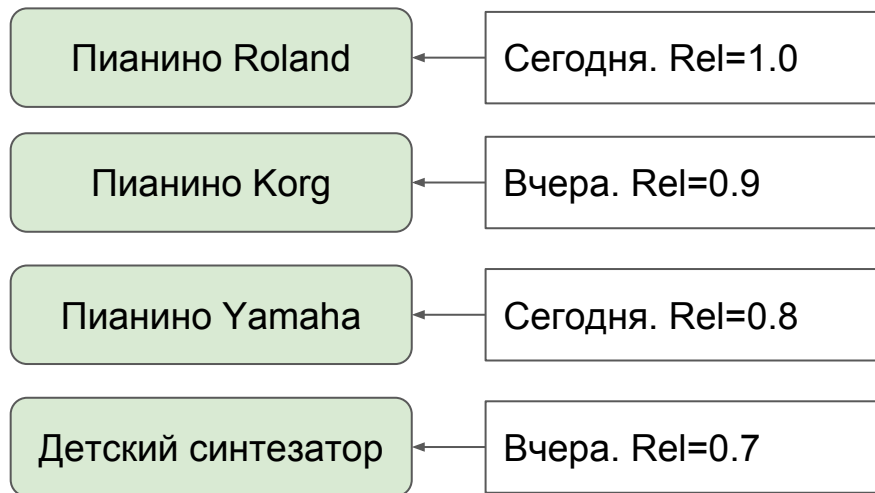
Money



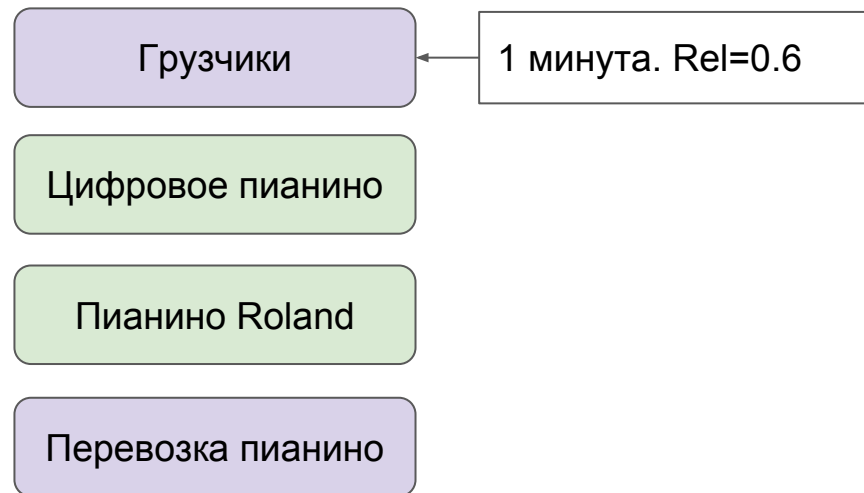
Общая проблема

Любая категория ▾	пианино	По всей России ▾	Найти
-------------------	---------	------------------	-------

Buyer



Money



Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Сегодня. Rel=1.0

Пианино Korg

Вчера. Rel=0.9

Пианино Yamaha

Сегодня. Rel=0.8

Детский синтезатор

Вчера. Rel=0.7

Money

Грузчики

1 минута. Rel=0.6

Цифровое пианино

10 минут. Rel=0.8

Пианино Roland

Перевозка пианино

Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Сегодня. Rel=1.0

Пианино Korg

Вчера. Rel=0.9

Пианино Yamaha

Сегодня. Rel=0.8

Детский синтезатор

Вчера. Rel=0.7

Money

Грузчики

1 минута. Rel=0.6

Цифровое пианино

10 минут. Rel=0.8

Пианино Roland

30 минут. Rel=1.0

Перевозка пианино

Общая проблема

Любая категория



пианино

По всей России



Найти

Buyer

Пианино Roland

Сегодня. Rel=1.0

Пианино Korg

Вчера. Rel=0.9

Пианино Yamaha

Сегодня. Rel=0.8

Детский синтезатор

Вчера. Rel=0.7

Money

Грузчики

1 минута. Rel=0.6

Цифровое пианино

10 минут. Rel=0.8

Пианино Roland

30 минут. Rel=1.0

Перевозка пианино

50 минут. Rel=0.7

Общая проблема

Любая категория ▾	fender	По всей России ▾	Найти
-------------------	--------	------------------	-------

Buyer



Fender 1979

Fender Strat

Fender Telecaster

Гитары и усилители

1 минута. Rel=1.0

10 минут. Rel=0.8

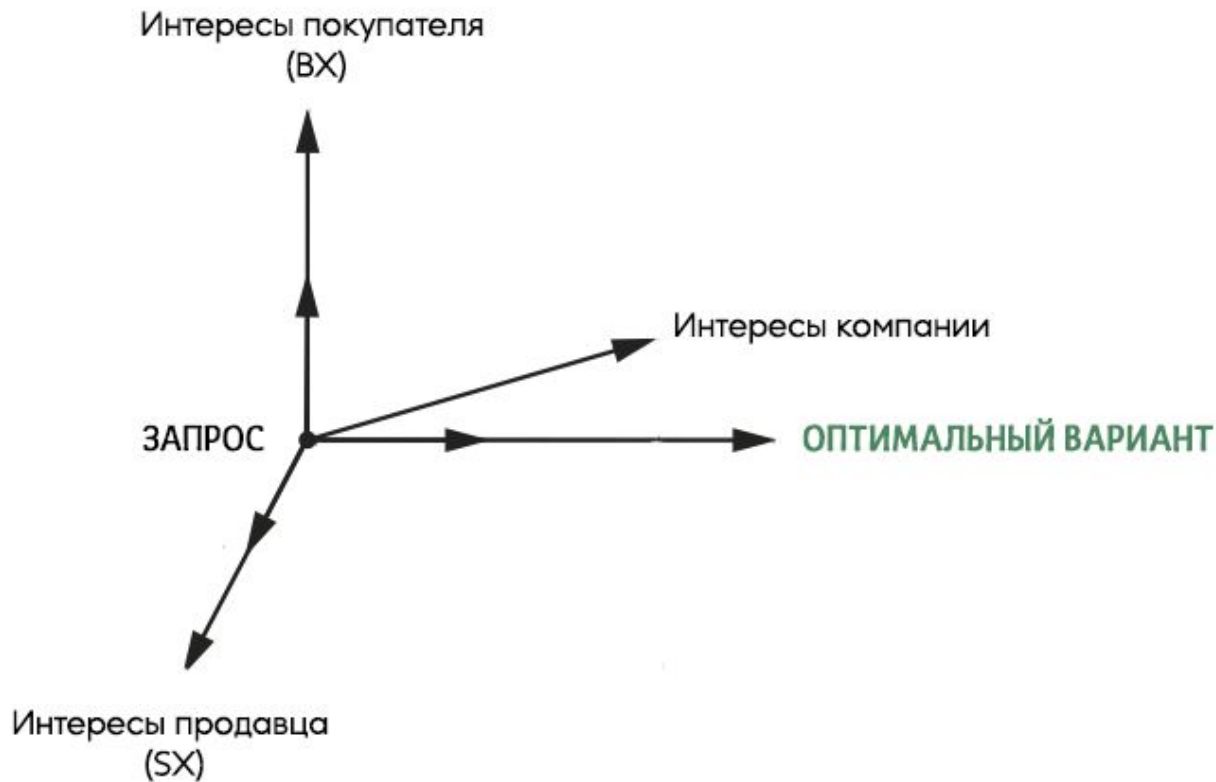
30 минут. Rel=1.0

50 минут. Rel=0.7

Money



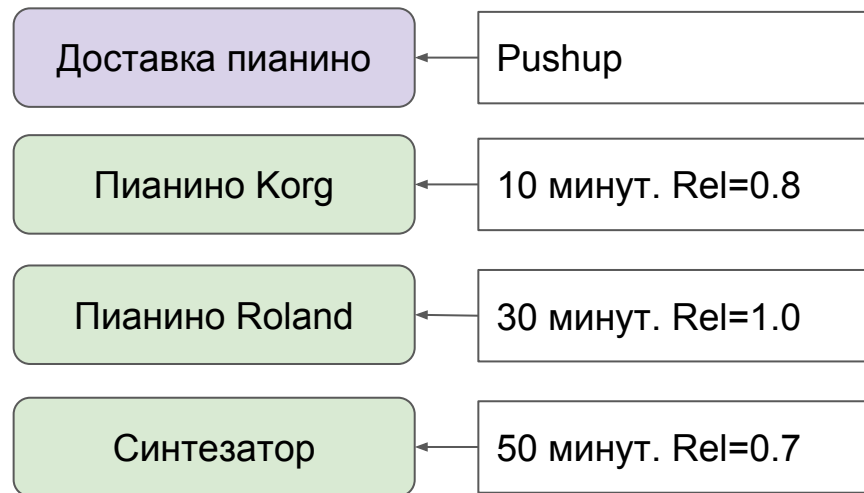
Что мы хотим сделать?



Возможные подходы

Вариант 1: Простое решение

2007 — 2017 В поиске Авито сортировка по времени



Вариант 1: Простое решение

Плюсы:

- Очень простое решение

Вариант 1: Простое решение

Плюсы:

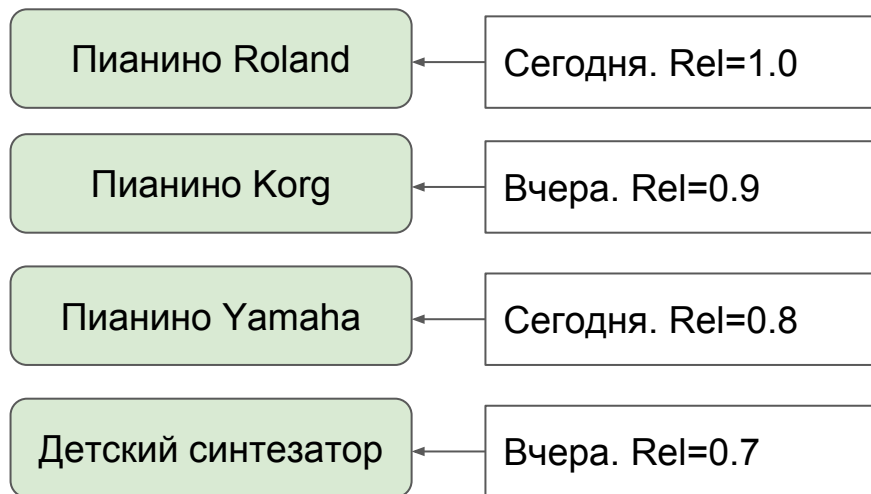
- Очень простое решение

Минусы:

- **Оптимизация невозможна**
- Сложно добавить релевантность
- Сложно добавлять новые услуги на выдаче

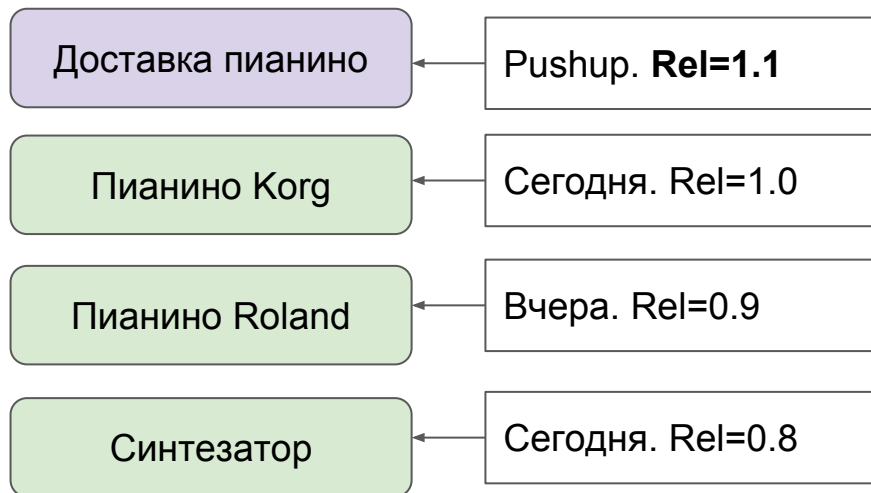
Вариант 2: Эвристика в ранжировании

2017 — 2018 Поиск стал релевантным



Вариант 2: Эвристика в ранжировании

2017 — 2018 Поиск стал релевантным



Вариант 2: Эвристика в ранжировании

- Появляется проблема Релевантность vs Монетизация

Плюсы:

- Возможность добавить релевантность в поиск
- **Оптимизация возможна**

Вариант 2: Эвристика в ранжировании

- Появляется проблема Релевантность vs Монетизация

Плюсы:

- Возможность добавить релевантность в поиск
- **Оптимизация возможна**

Минусы:

- **Слабая управляемость**
- Нельзя сильно менять формулу ранжирования
- Нельзя добавить машинное обучение
- Сложно добавлять новые услуги на выдаче

Вариант 3: Многокритериальная оптимизация

- **2018+** Необходимость разделить платные услуги и ранжирование

Требования:

- Независимость от алгоритма ранжирования
- Открывается возможность сильной оптимизации поиска, например ML
- Управляемая выдача платных
- Возможность добавлять любые другие услуги (доставка запчастей)

Вариант 3: Многокритериальная оптимизация

- **2018+** Необходимость разделить платные услуги и ранжирование

Требования:

- Независимость от алгоритма ранжирования
- Открывается возможность сильной оптимизации поиска, например ML
- Управляемая выдача платных
- Возможность добавлять любые другие услуги (доставка запчастей)

Проблемы:

- Более сложная реализация
- Сложнее интегрировать в существующую систему

Блендер

Интенст запроса

Любая категория	▼	ягуар	По всей России	▼	Найти
-----------------	---	-------	----------------	---	-------



Jaguar XJ, 2017

5 100 000 ₽

445 км, 3.0 AT (340 л.с.), седан, полный, бензин

[Проверить историю автомобиля](#)

[Jaguar Land Rover Лаки Моторс](#)

Екатеринбург

1 час назад

User 1



User 2



Интенст запроса

Любая категория	▼	ягуар	По всей России	▼	Найти
-----------------	---	-------	----------------	---	-------



Jaguar XJ, 2017

5 100 000 ₽

445 км, 3.0 AT (340 л.с.), седан, полный, бензин

[Проверить историю автомобиля](#)

[Jaguar Land Rover Лаки Моторс](#)

Екатеринбург

1 час назад

User 1



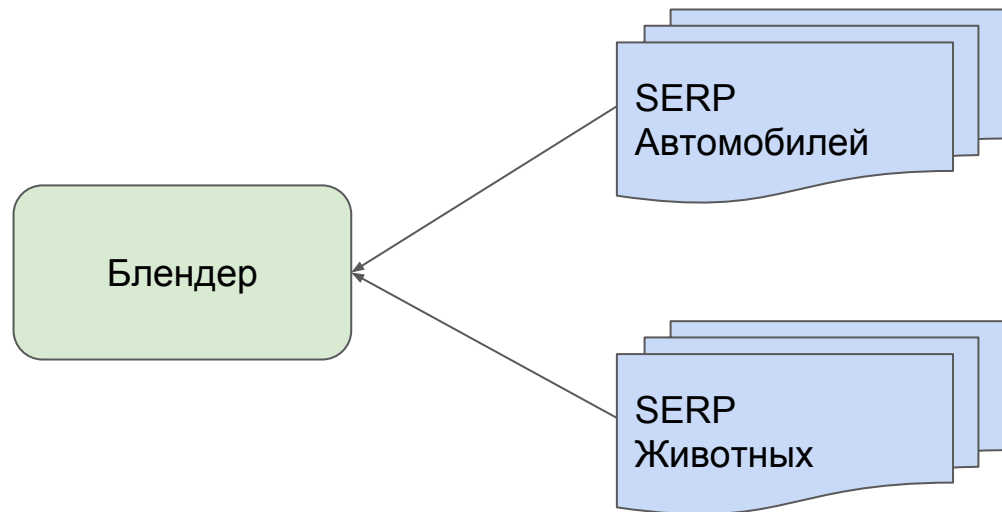
0.9

0.1

User 2



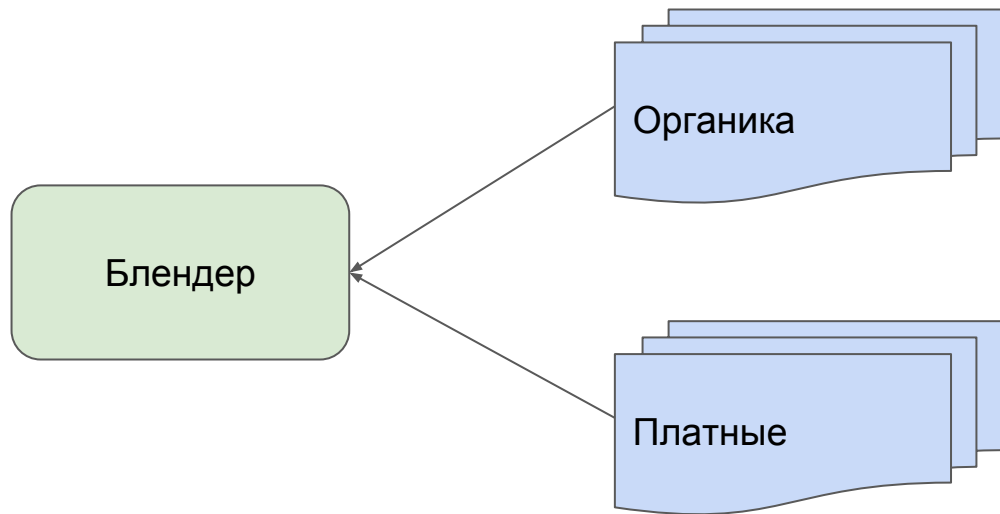
Интенст запроса



Интенст запроса

Объявление	Тип	Любители машин	Любители животных
Ручной ягуар		0	0.9
Jaguar JX 2017		0.9	0
Домашний ягуар		0	0.8
Jaguar JX 2016		0.8	0

Интенд запроса



Как мы смотрим выдачу?


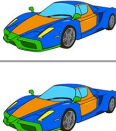

Пример

Просмотр

Успех

Не нашли



ID	Тип	Текст
1		Новый Jaguar
2		Jaguar с пробегом
3		Jaguar 2017
4		Jaguar 2016




Пример

Просмотр

Успех

Не нашли



ID	Тип	Текст
1		Новый Jaguar
2	 	Jaguar с пробегом
3		Jaguar 2017
4		Jaguar 2016



Пример

Просмотр

Успех

Не нашли



ID	Тип	Текст
1		Новый Jaguar
2		Jaguar с пробегом
3	 	Jaguar 2017
4		Jaguar 2016



Пример

Просмотр

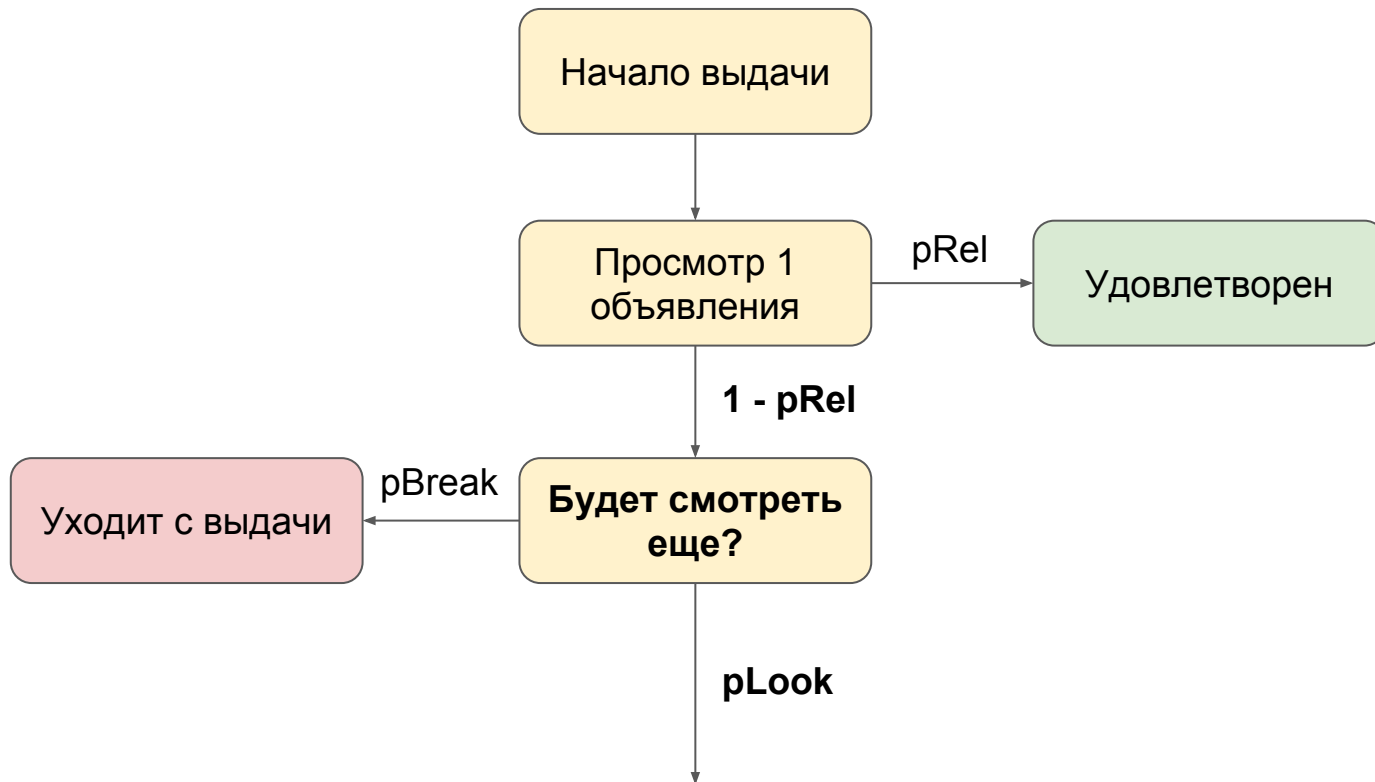
Успех

Не нашли

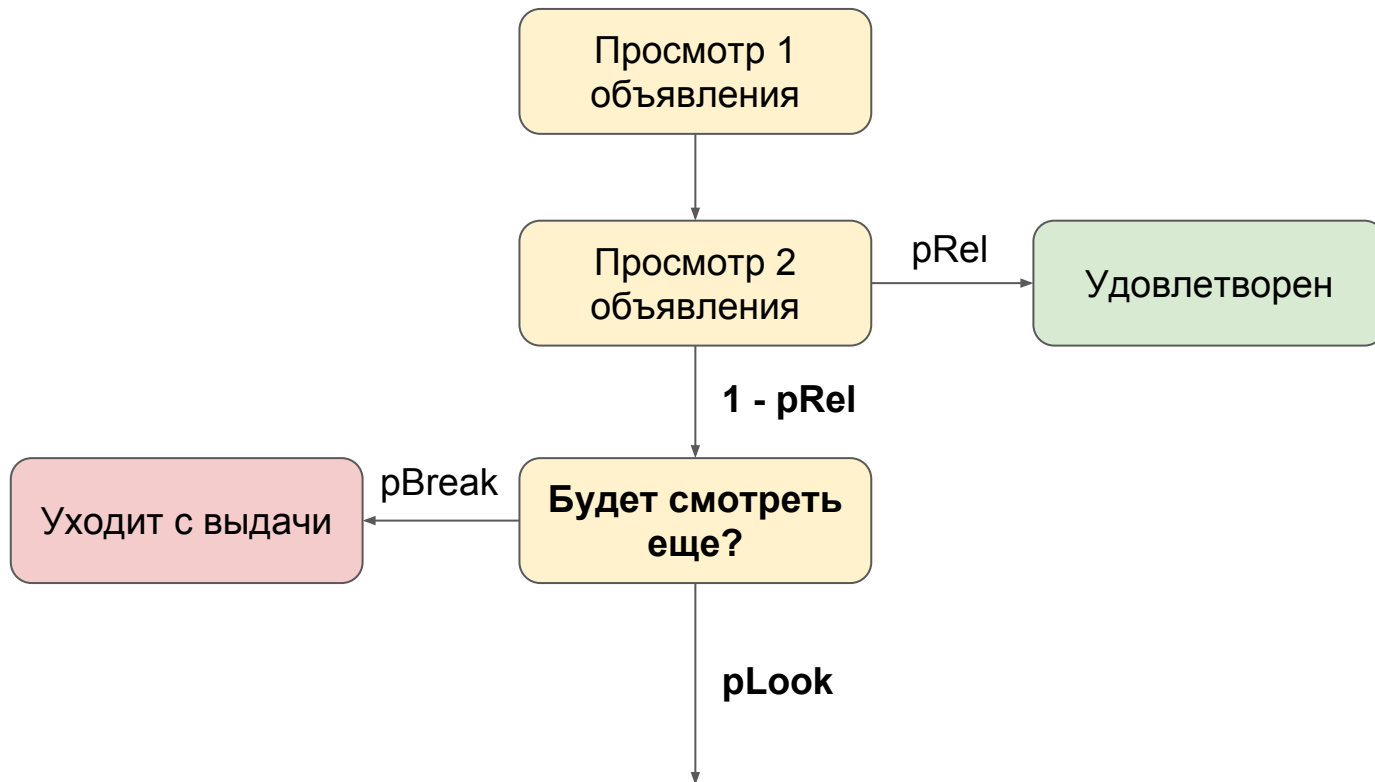


ID	Тип	Текст
1		Новый Jaguar
2		Jaguar с пробегом
3	 	Jaguar 2017
4		Jaguar 2016

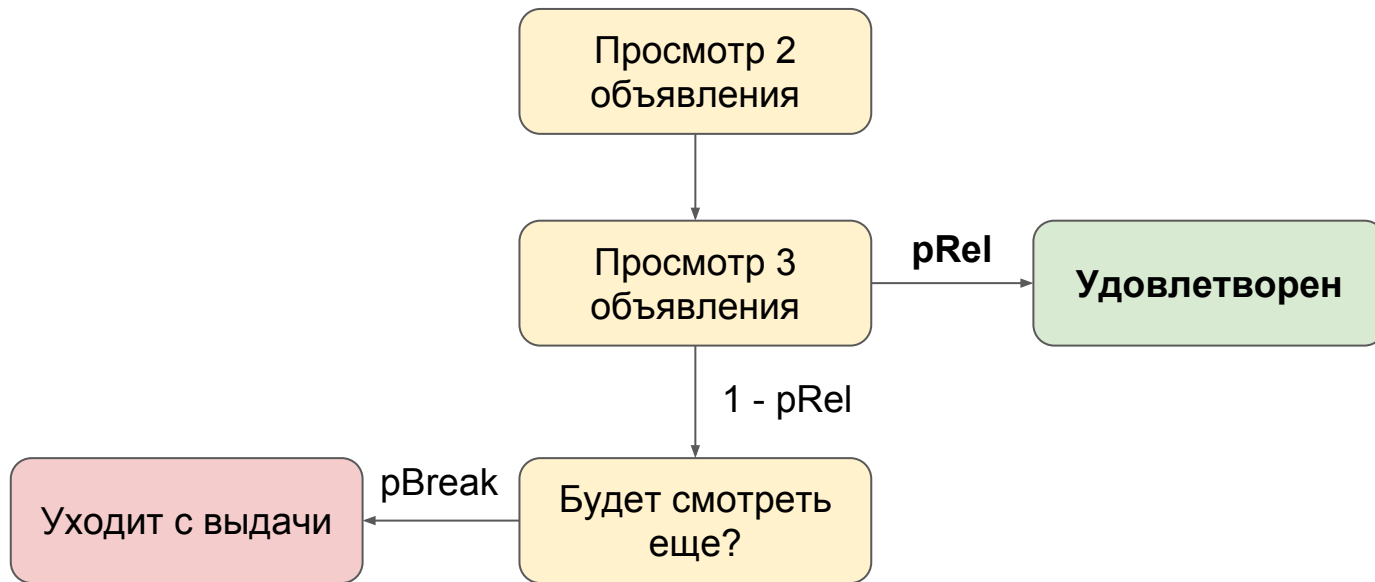
Вероятность просмотра



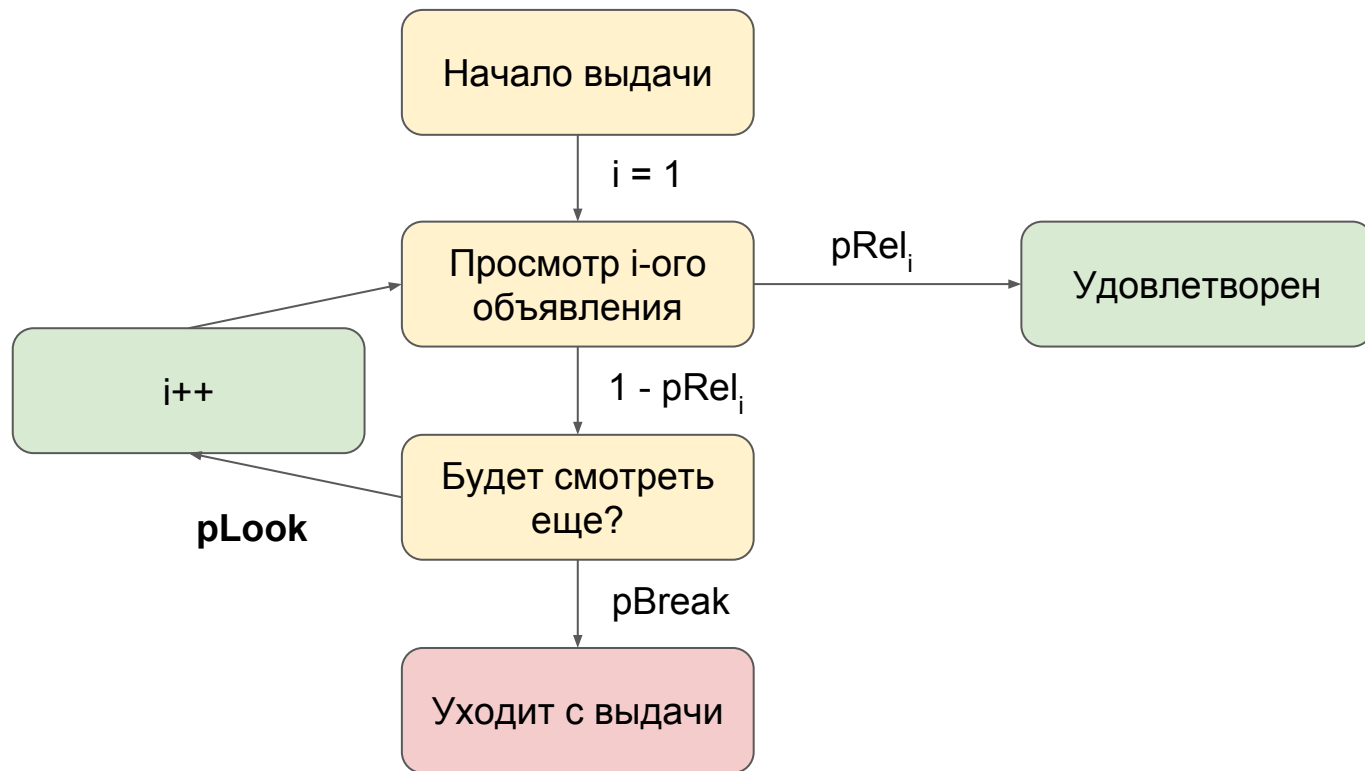
Вероятность просмотра



Вероятность просмотра



Вероятность просмотра



Блендер

Lookers

Lucky ones

Breakers

Обычная выдача

Пример

Lookers



Lucky ones

Breakers

Объявление	Тип	Любители машин	Любители животных
1 		0.9	0
2		0.8	0
3		0.7	0
4		0.6	0

Пример

Lookers



Lucky ones

Breakers

Объявление	Тип	Любители машин	Любители животных
1		0.9	0
2	 	0.8	0
3		0.7	0
4		0.6	0

Пример

Lookers



Lucky ones

Breakers



Объявление	Тип	Любители машин	Любители животных
1		0.9	0
2 		0.8	0
3		0.7	0
4		0.6	0

Пример

Lookers



Lucky ones

Breakers



Объявление	Тип	Любители машин	Любители животных
1		0.9	0
2		0.8	0
3	 	0.7	0
4		0.6	0

Пример

Lookers

Lucky ones

Breakers



Объявление	Тип	Любители машин	Любители животных
1		0.9	0
2		0.8	0
3		0.7	0
4		0.6	0

Смешанная выдача

Блендер

Lookers



Lucky ones

Breakers

Объявление	Тип	Любители машин	Любители животных
1 		0	0.9
2		0.9	0
3		0	0.8
4		0.8	0

Блендер

Lookers



Lucky ones



Breakers

Объявление	Тип	Любители машин	Любители животных
1 		0	0.9
2		0.9	0
3		0	0.8
4		0.8	0

Блендер

Lookers



Lucky ones



Breakers

Объявление	Тип	Любители машин	Любители животных
1		0	0.9
2 		0.9	0
3		0	0.8
4		0.8	0

Блендер

Lookers



Lucky ones



Breakers

Объявление	Тип	Любители машин	Любители животных
1		0	0.9
2		0.9	0
3		0	0.8
4		0.8	0

Блендер

Lookers



Lucky ones



Breakers

Объявление	Тип	Любители машин	Любители животных
1		0	0.9
2		0.9	0
3		0	0.8
4	 	0.8	0

Блендер

Lookers

Lucky ones

Breakers



Объявление	Тип	Любители машин	Любители животных
1		0	0.9
2		0.9	0
3		0	0.8
4		0.8	0

Вероятность просмотра

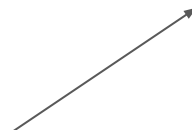
Качество выдачи

$$p_{found} = \sum_{i=1}^n pLook[i] * pRel[i]$$

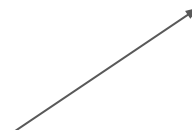
Вероятность
просмотра

$$pLook[i] = pLook[i-1] * (1 - pRel[i-1]) * (1 - pBreak)$$

не удовлетворен
просмотренным

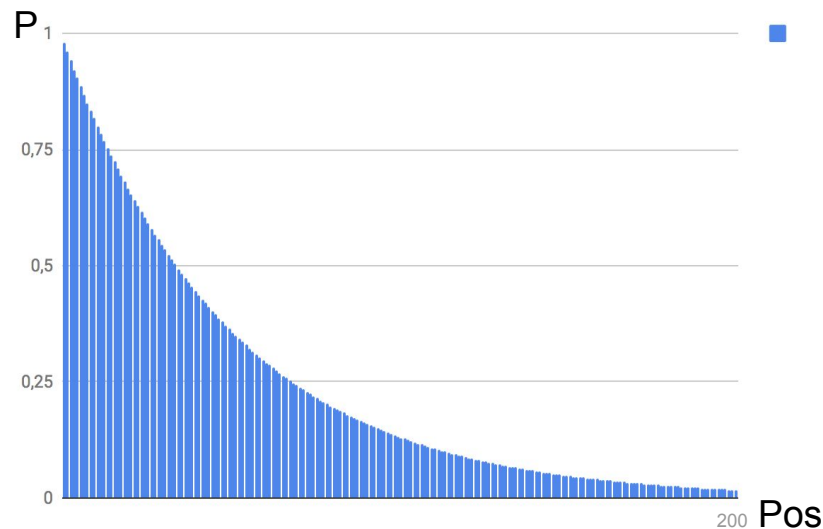


не ушел с
выдачи

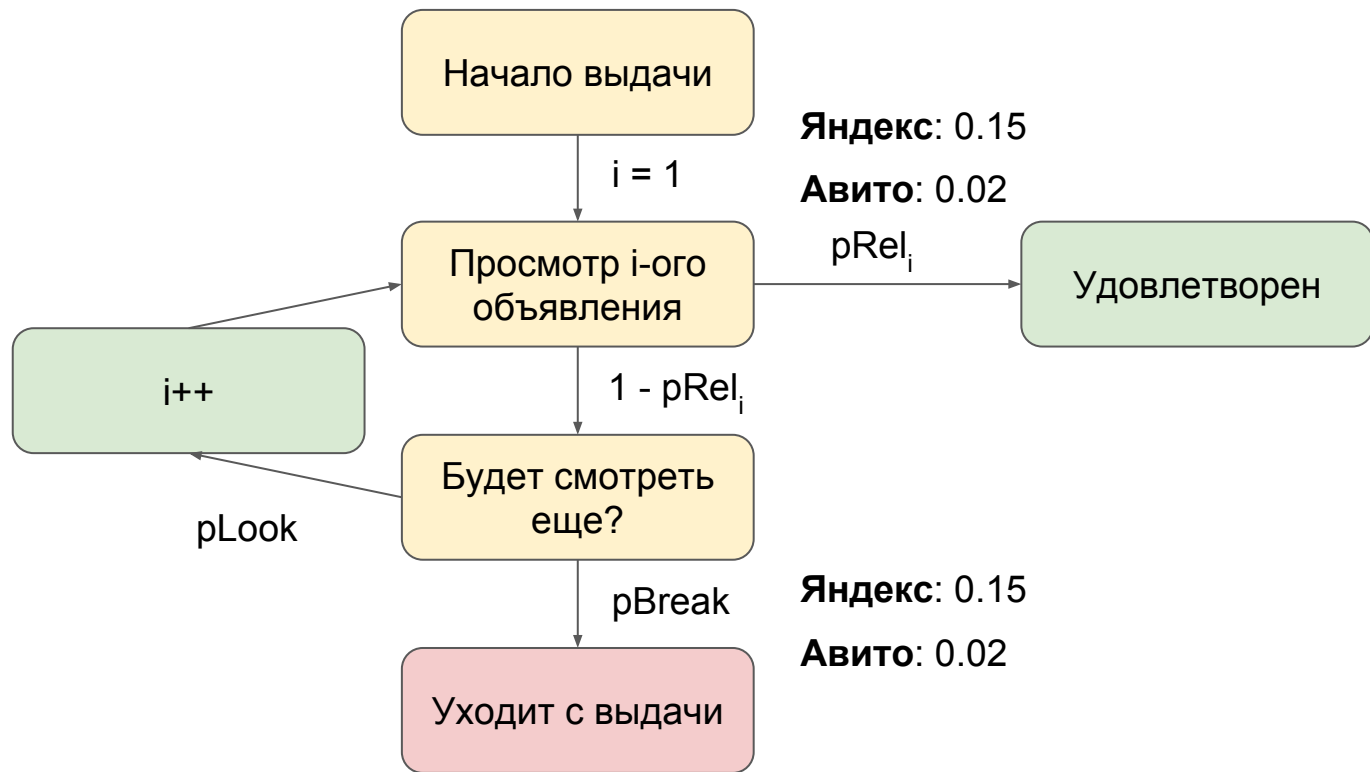


Вероятность просмотра

$$p\text{Look} = 0.98^{\text{pos}}$$



Вероятность просмотра



Жадный алгоритм

1. Получаем органическую и платную выдачи
2. Определяем веса интересов (вероятности интентов)
3. Заполняем таблицу релевантностей объявлений интентам
4. На каждом шаге берем N объявлений из каждой выдачи
5. Находим объявление с максимальным скалярным произведением векторов текущих интентов и релевантности интентам
6. Обновляем состояние блендера
7. Если еще есть объявления, переходим к п. 4

Начальное состояние

Соотношение, в котором мы должны удовлетворить разные интересы на данном серпе



Пример смешивания

Получаем органическую и платную выдачи

Хорошие объявления

Fender Start

Fender Telecaster

Хорошие платные

Fender 1979

Fender Custom

Top 50 платных

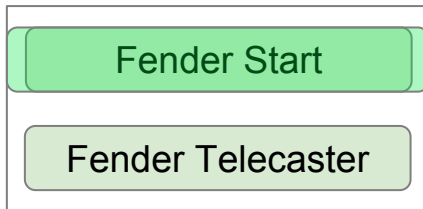
Обучение гитаре

Усилители

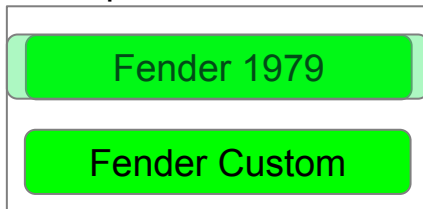
Выдача

На каждом шаге берем N объявлений из каждой выдачи

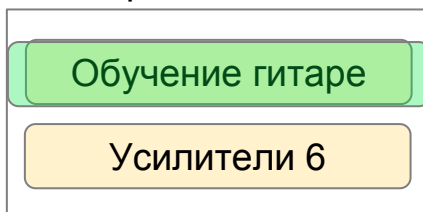
Хорошие объявления



Хорошие платные



Top 50 платных



Выдача



На каждом шаге берем N объявлений из каждой выдачи

Хорошие объявления

Fender Telecaster

Хорошие платные

Fender Custom

Top 50 платных

Усилители

Fender Start

Fender 1979

Обучение гитаре

Выдача

Шаг 2. Выбираем лучшее

Находим объявление с максимальным скалярным произведением

Хорошие объявления

Fender Telecaster

Хорошие платные

Fender Custom

Top 50 платных

Усилители

Fender Start

Fender 1979

Обучение гитаре

Выдача

Находим объявление с максимальным скалярным произведением

Хорошие объявления

Fender Start

Fender Telecaster

Хорошие платные

Fender Custom

Топ 50 платных

Обучение гитаре

Усилители

Выдача

Fender 1979

Шаг 3. Расчет счастья

Хорошие объявления

Fender Start

Fender Telecaster

Хорошие платные

Fender Custom

Top 50 платных

Обучение гитаре

Усилители

Money: 0.1
Buyer: 0.1

Выдача

Fender 1979

Хорошие объявления

Fender Telecaster

Хорошие платные

Fender Custom

Топ 50 платных

Обучение гитаре

Усилители

Money: 0.1
Buyer: 0.2

Выдача

Fender 1979

Fender Start

Хорошие объявления

Fender Telecaster

Хорошие платные

Fender Custom

Top 50 платных

Усилители

Money: 0.4
Buyer: 0.2

Выдача

Fender 1979

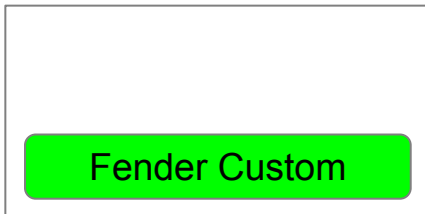
Fender Start

Обучение гитаре

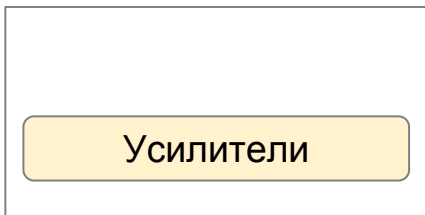
Хорошие объявления



Хорошие платные



Top 50 платных



Money: 0.4
Buyer: 0.3

Выдача



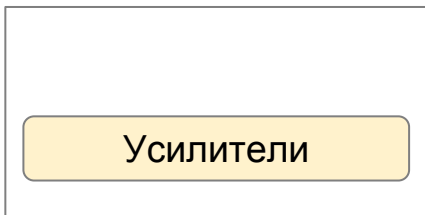
Хорошие объявления



Хорошие платные



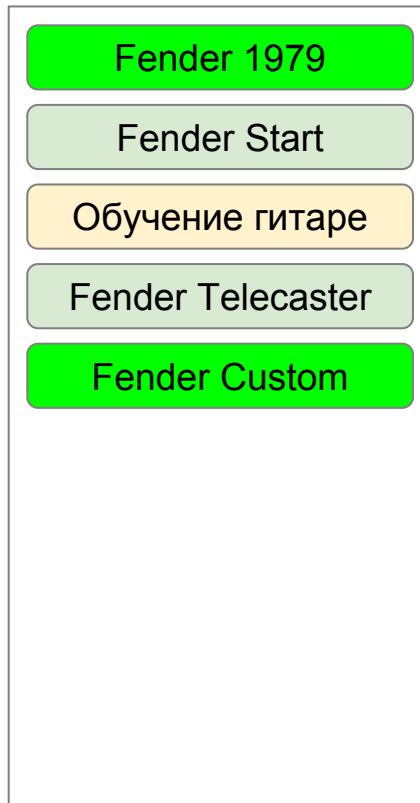
Top 50 платных



Усилители

Money: 0.5
Buyer: 0.5

Выдача



Fender 1979

Fender Start

Обучение гитаре

Fender Telecaster

Fender Custom

Детали

Заполнение таблицы

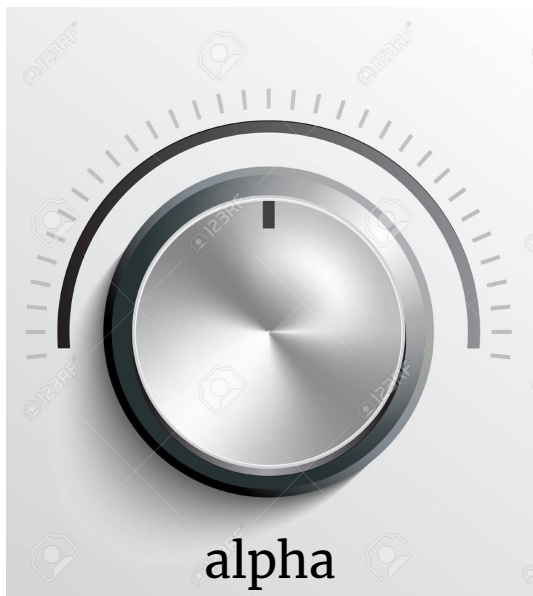
$$\left\{ \begin{array}{l} \text{Buyer} = k * pos + b \\ \text{Money} = \text{norm}(0.98 ^ \wedge age) \end{array} \right.$$

$$\left\{ \begin{array}{l} \text{Buyer} = \text{norm}(\text{rank}) \\ \text{Money} = \text{norm}(0.98 ^ \wedge age) \end{array} \right.$$

Максимизируем интересы всех сторон

Начальное состояние

- *Buyer = 1 - alpha*
- *Money = alpha*



$B = 0.7$ vs $M = 0.3$

$B = 0.7$ vs $M = 0.3$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино раритет

Пианино БУ

Топ 50 платных

Грузчики

Перевозки

Money: 0.0
Buyer: 0.0

Выдача

$B = 0.7$ vs $M = 0.3$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Грузчики

Перевозки

Money: 0.1
Buyer: 0.2

Выдача

Пианино раритет

$B = 0.7$ vs $M = 0.3$

Хорошие объявления

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Грузчики

Перевозки

Money: 0.1
Buyer: 0.4

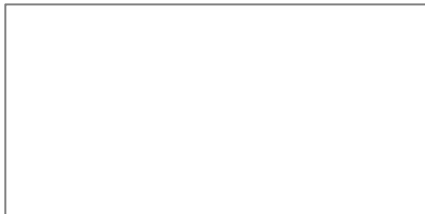
Выдача

Пианино раритет

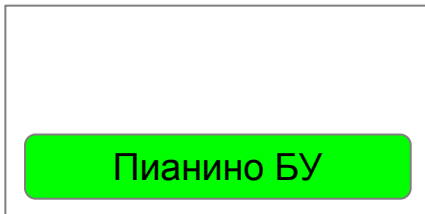
Пианино Korg

$B = 0.7$ vs $M = 0.3$

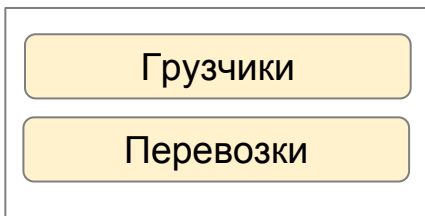
Хорошие объявления



Хорошие платные

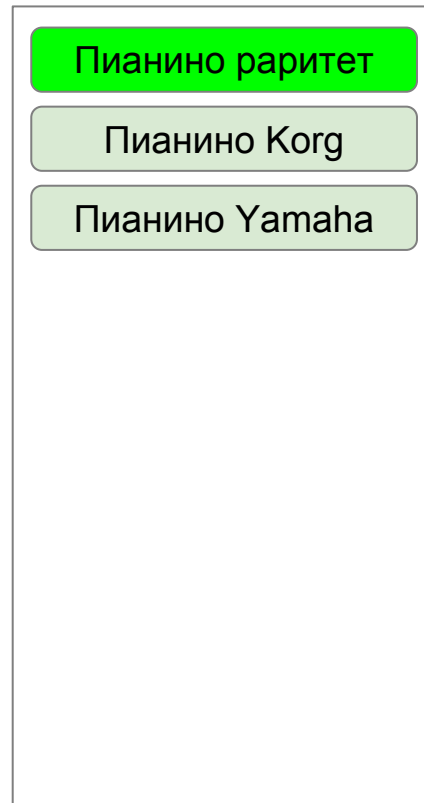


Топ 50 платных



Money: 0.1
Buyer: 0.6

Выдача



$B = 0.7$ vs $M = 0.3$

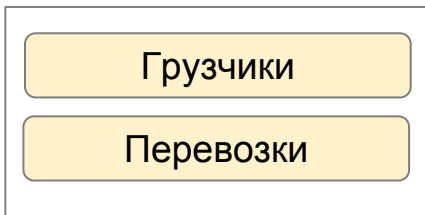
Хорошие объявления



Хорошие платные

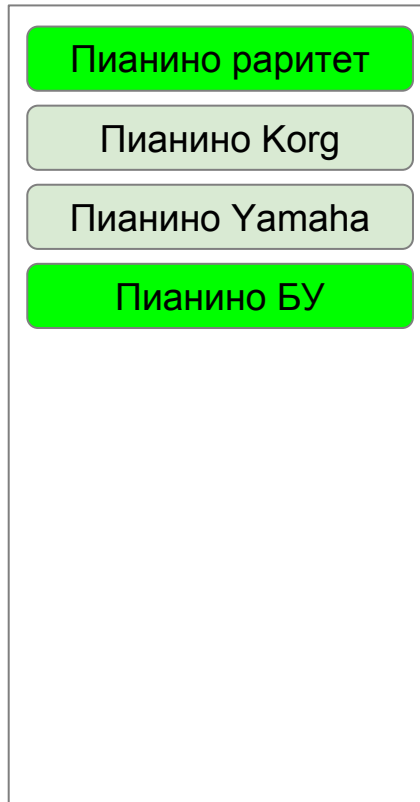


Топ 50 платных



Money: 0.2
Buyer: 0.7

Выдача



$B = 0.7$ vs $M = 0.3$

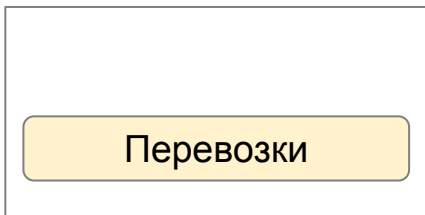
Хорошие объявления



Хорошие платные



Топ 50 платных



Money: 0.3
Buyer: 0.7

Выдача



$B = 0.1$ vs $M = 0.9$

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино раритет

Пианино БУ

Топ 50 платных

Грузчики

Перевозки

Money: 0.0
Buyer: 0.0

Выдача

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Грузчики

Перевозки

Money: 0.1
Buyer: 0.1

Выдача

Пианино раритет

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Перевозки

Money: 0.2
Buyer: 0.1

Выдача

Пианино раритет

Грузчики

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Money: 0.5
Buyer: 0.1

Выдача

Пианино раритет

Грузчики

Перевозки

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Money: 0.9
Buyer: 0.1

Выдача

Пианино раритет

Грузчики

Перевозки

Грузчики

Грузчики

Грузчики

Грузчики

Не все так просто

0.1 vs 0.9 протекание

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино раритет

Пианино БУ

Топ 50 платных

Грузчики

Перевозки

Money: 0.0
Buyer: 0.0

Выдача

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Грузчики

Перевозки

Money: 0.1
Buyer: 0.1

Выдача

Пианино раритет

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Перевозки

Money: 0.2
Buyer: 0.1

Выдача

Пианино раритет

Грузчики

$B = 0.1$ vs $M = 0.9$

Хорошие объявления

Пианино Korg

Пианино Yamaha

Хорошие платные

Пианино БУ

Топ 50 платных

Money: 0.5
Buyer: 0.1

Выдача

Пианино раритет

Грузчики

Перевозки

$B = 0.1$ vs $M = 0.9$

Хорошие объявления



Хорошие платные



Топ 50 платных



Money: 0.6

Buyer: 0.4

Выдача

Пианино раритет

Грузчики

Перевозки

Грузчики

Пианино Korg

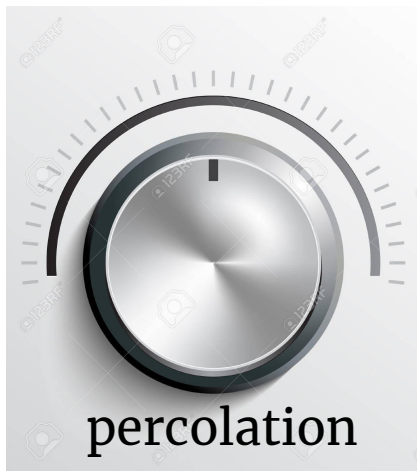
Пианино Yamaha

Пианино БУ

ошибка накапливается и уравнивается

Протекание

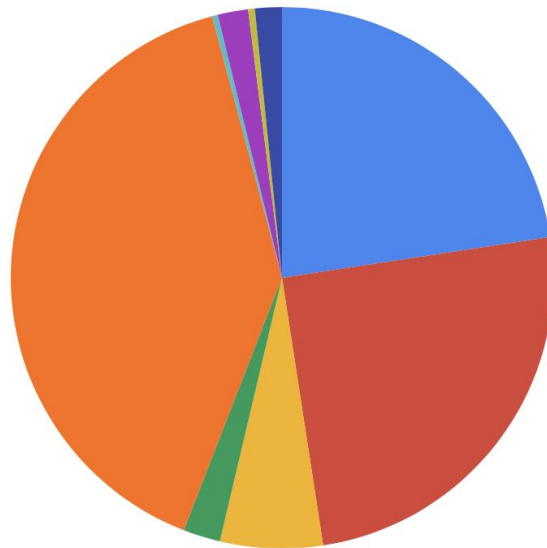
$$\text{State} = (1-P) * \text{NewState} + P * \text{StartState}$$



Начальное состояние

- *Buyer = 1 - alpha*
- *Money = alpha*

Top Secret



Начальное состояние

- $K = f(\text{SERP}, \text{total_found})$
- $BX = K / (K + \text{alpha})$
- $SX = \text{alpha} / (K + \text{alpha})$



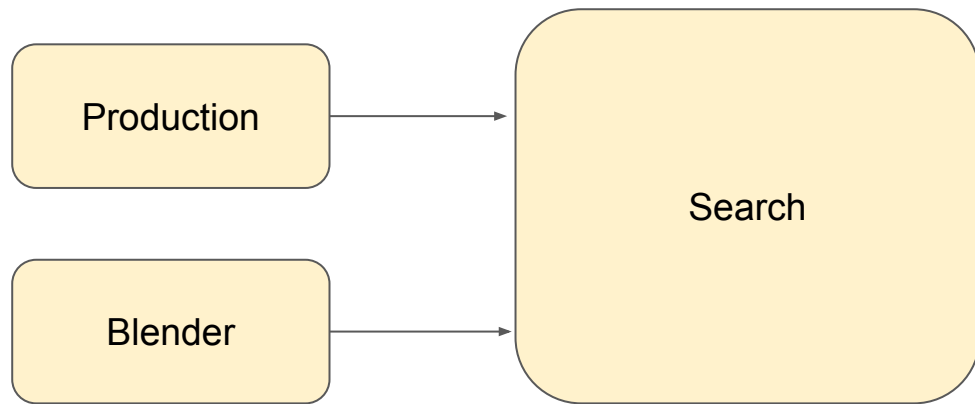
Как это выкатить в production?

Как это выкатить в production?

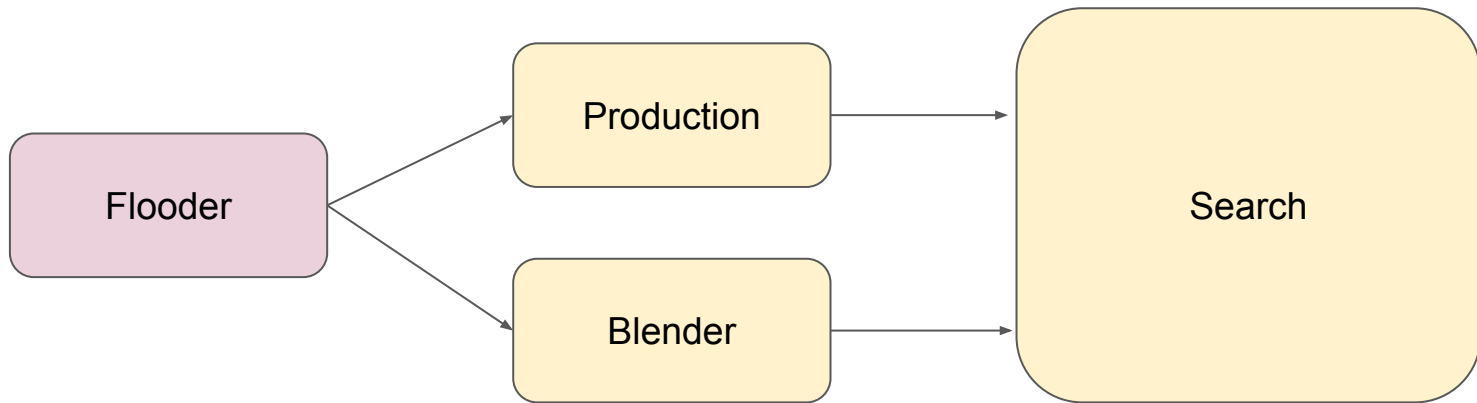
Цель MVP:

- Воспроизвести эвристику через блендер
- Доказать возможность контроля платных на выдаче
- Проверить нагрузку
- Минимально затрагивать пользователей и монетизацию во время АБ-тестов

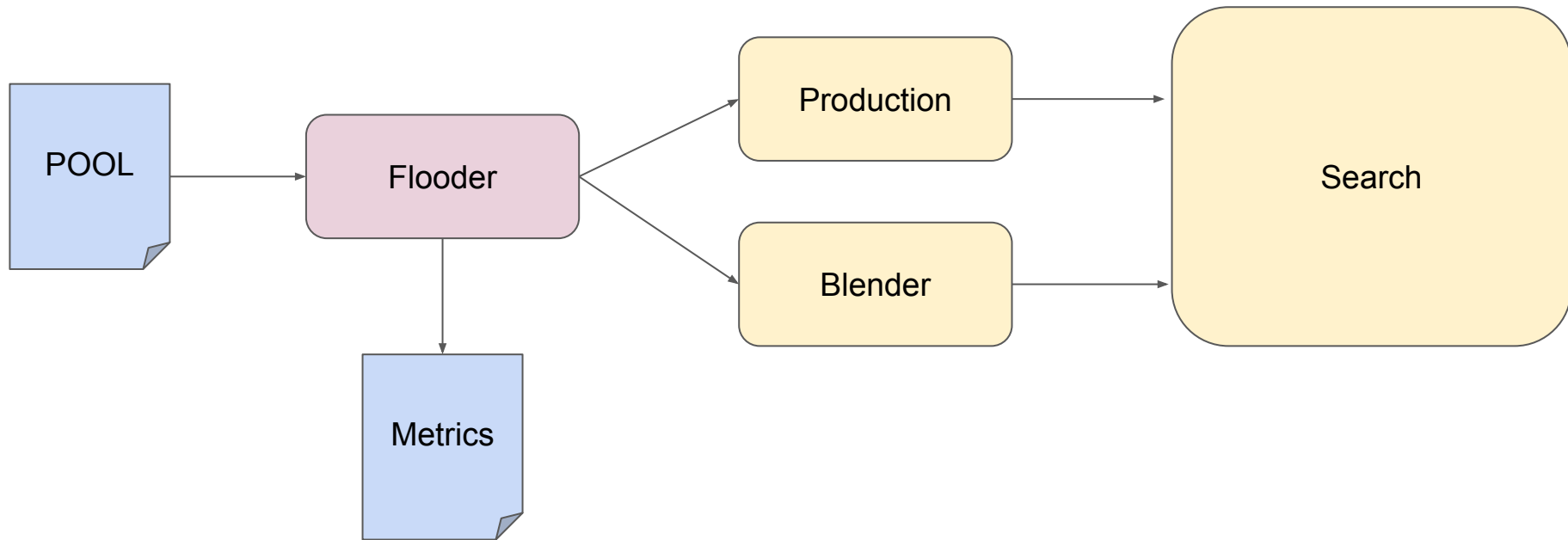
Как это выкатить в production?



Как это выкатить в production?



Как это выкатить в production?



Как это выкатить в production?

- Срез по страницам (берем top-200, делим на 50)
- Срез по категориям (разделяем по cat_id)
- Срез по дням (свежесть в днях)

Как это выкатить в production?

- `push imp` — количество платных
- `items imp` — количество объявлений
- `imp item age` — средний возраст объявления
- `delivery imp` — количество смешанной доставки
- здравый смысл (история про разнообразие)

`sum([fun(elem) * 0.98**elem.pos for elem in serp]) / len(queries)`

Подводные камни

page	push_imp old	push_imp blender
1	10,266	10,536
2	0,984	3,883
3	0,125	1,716
4	0,019	0,288

Подводные камни

page	push_imp old	push_imp blender
1	10,266	9,136
2	0,984	1,383
3	0,125	0,516
4	0,019	0,188

Подводные камни

page	push_imp old	push_imp blender
1	10,266	10,536
2	0,984	3,883
3	0,125	1,716
4	0,019	0,288
total	11,394	16,423

page	push_imp old	push_imp blender
1	10,266	9,136
2	0,984	1,383
3	0,125	0,516
4	0,019	0,188
total	11,394	11,223

Запуски и АБ тесты

Здравый смысл



Раки

370 ₽

Продукты питания | Компания
Новороссийск
1 день назад



Раки

900 ₽

Продукты питания
Волоколамск
18 часов назад



Раки

500 ₽

Продукты питания
Мичуринск
2 дня назад



Щебень

830 ₽

Ремонт и строительство | Компания
Новосибирск
5 часов назад



Щебень

500 ₽

Ремонт и строительство | Компания
Волгоград
6 часов назад



Щебень

600 ₽

Ремонт и строительство | Компания
Волгоград
6 часов назад

Здравый смысл

Выводы

- Добавили метрику разнообразия выдачи
- Всегда делаем side-by-side-приемку ранкеров

Результаты экспериментов

Первый запуск

- Пул по репрезентативному потоку
- Offline-ошибка $\pm 5\%$
- Трафик 20/20 %
- Стат значимый замер push impression
- Результат: **+25,6%** платных :(

Результаты экспериментов

Второй запуск

- Обновили процесс настройки
- Пул по репрезентативному потоку
- Offline-ошибка $\pm 3\%$
- Трафик 20/20 %
- Стат значимый замер push impression
- Результат: **+14,7%** платных

Результаты экспериментов

Третий запуск

- Собрали отдельно пулы в разных регионах
- Поток + регионы
- Offline-ошибка $\pm 3\%$
- Трафик 20/20 %
- Стат значимый замер push impression
- Результат: **+1.6%** платных

Дополнительный профит

Быстрое добавление новых продуктов

Чеклист

- Добавить формулу нового интента
- Пробросить необходимые флаги в базовый поиск
- Добавить вес нового интента в блендер
- Добавить новый серп к смешиванию и условие его использования
- Обновить настройки метапоиска
- Приемка, АБ-тест, деплой

Запуск доставки запчастей

- Отдельный пул для запчастей в разных регионах
- Трафик 20/20 %
- Стат значимый замер item views serp
- Результат: **+3.4%** просмотров, выдачи в регионах стали больше

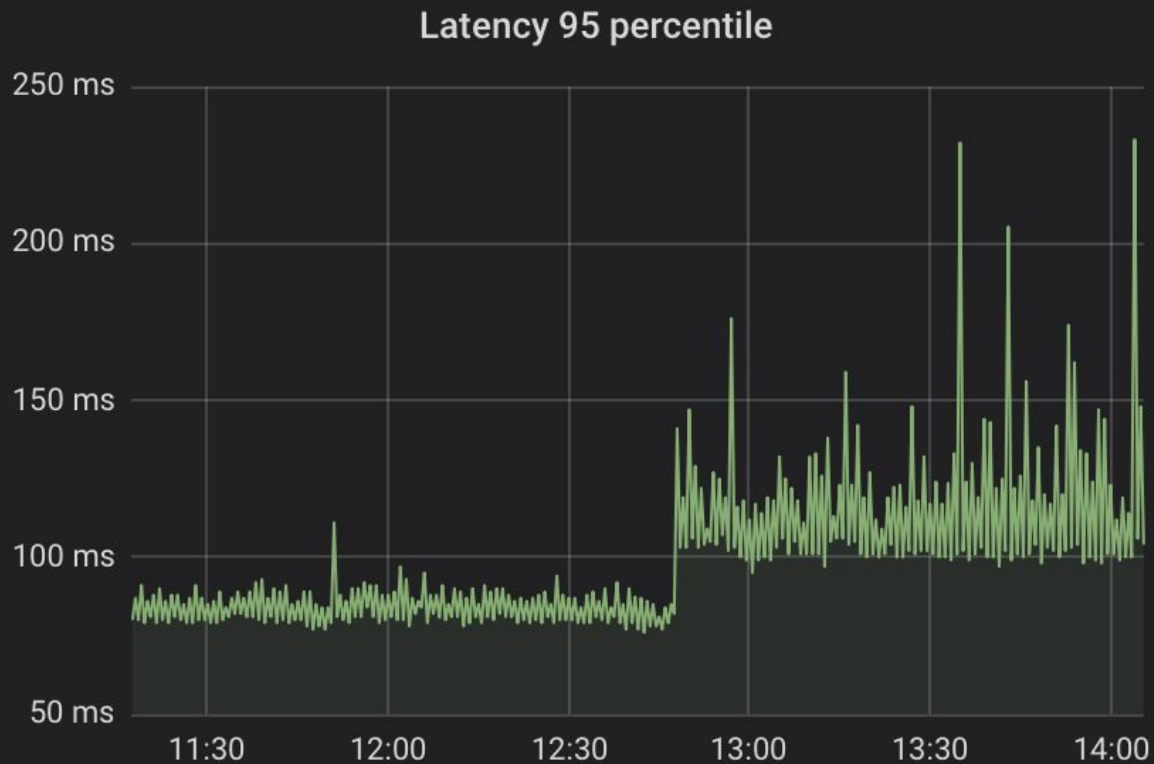


delivery force



delivery serp limit

Инфраструктура и нагрузка



Обменяли **20% CPU** на контроль платных и возможность
улучшения качества поиска

Выводы. Качество

- **Независимая оптимизация**
- Приводим все к вероятностям
- **Оффлайн настройка на модели с пулом**
- Простые формулы — осязаемый результат
- Не только метрики, но и глазами

Выводы. Инфраструктура

- Параллельные запросы с разными интентами
- Утилиты для анализа качества
- Утилиты для сбора пулов
- Возможность гибко настраивать алгоритмы
- Готовность к трейдоффам

Профиты

- Открыли путь к ранжированию на ML
- Контроль платных и новые возможности
- Легкие запуски продуктов
- Подмешивание запчастей с доставкой
- С 2018 года блендер в production