

Item2item service

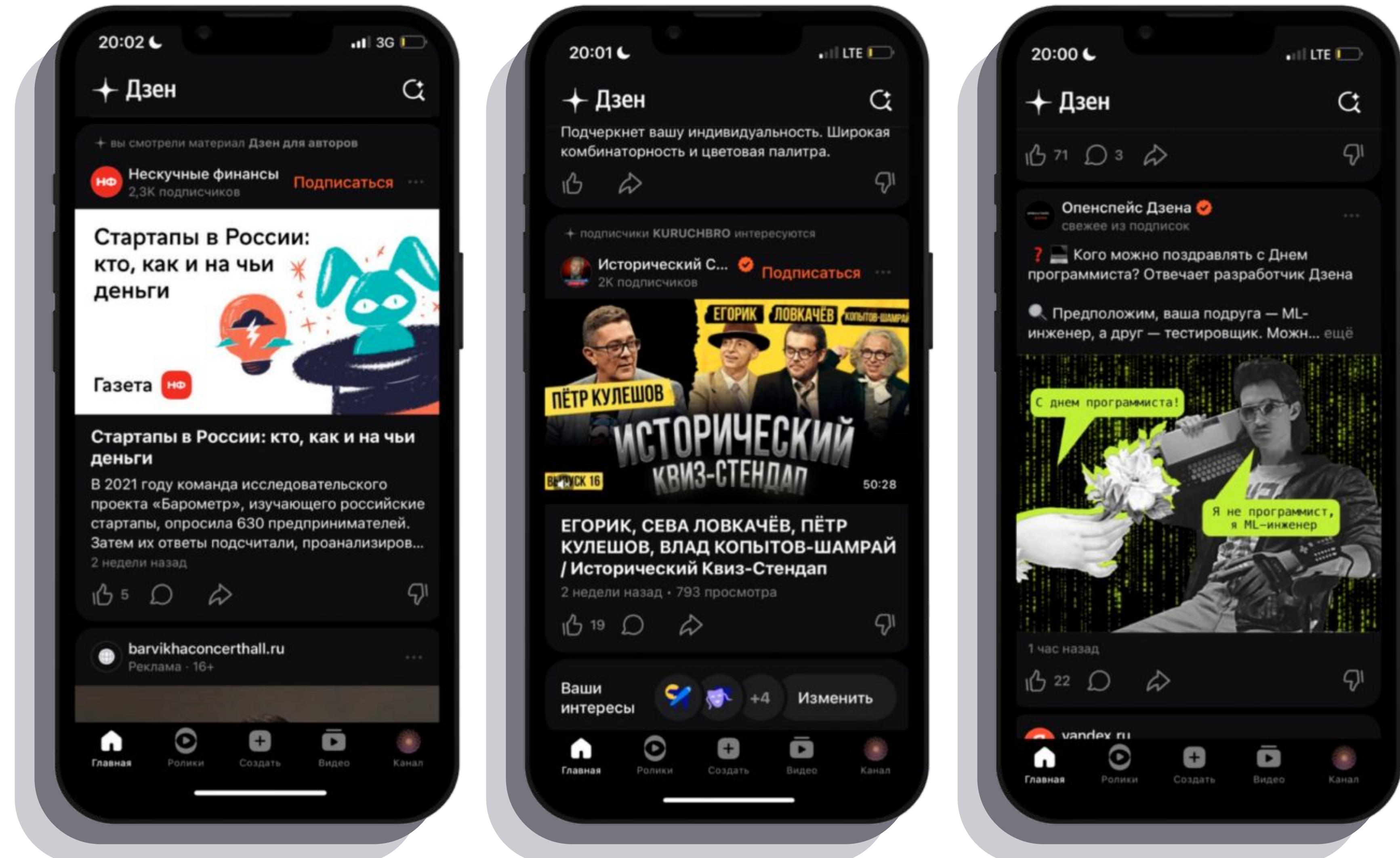
Vladislav Goncharenko

Slides by Dima Shishov, Dzen



MSU, fall 2023

Дзен



Дзен

Saint
HighLoad++
2023



Дзен в цифрах

DAU 30M+ пользователей

MAU 70M+ пользователей

35 минут в день

10K RPS на рекомендер

100M+ пользователей с историей

5 млн документов в базе
рекомендаций

10 млрд событий в день

500мс на ответ



Saint
HighLoad++
2023



О чём мы поговорим

- ★ Метрики Дзена и найденные проблемы
- ★ item2item схема рекомендаций
- ★ Рекомендер Дзена и i2i-сервис в нём
- ★ Результаты

Метрики Дзена и найденные проблемы



Top-level метрики сервиса

DAU

NPS

Timespent



Saint
HighLoad++
2023



Top-level метрики сервиса

Как растить? DAU



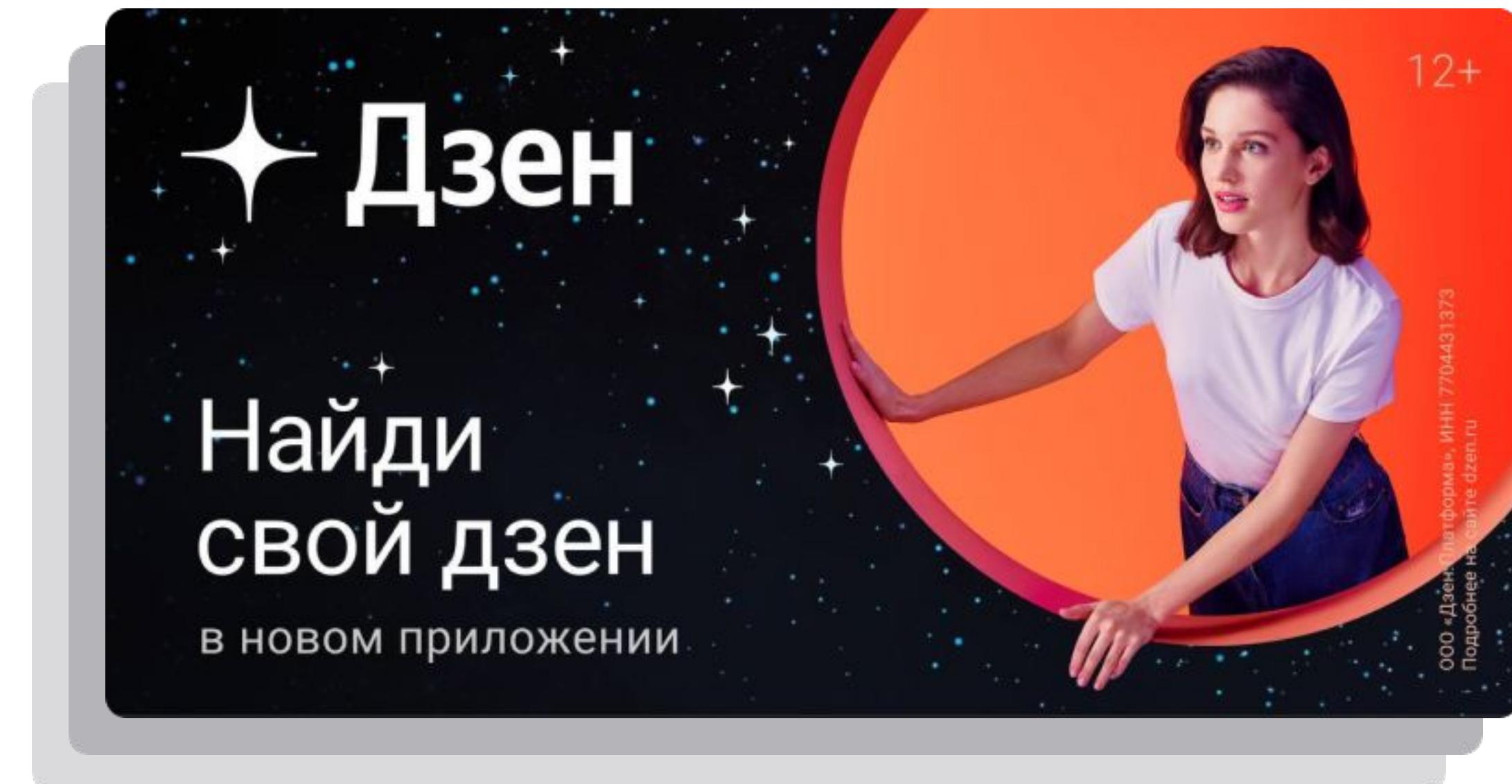
Saint
HighLoad++
2023



Top-level метрики сервиса

Как растить? DAU

- ❖ Маркетинговыми кампаниями
- ❖ Эксклюзивным контентом
- ❖ SEO



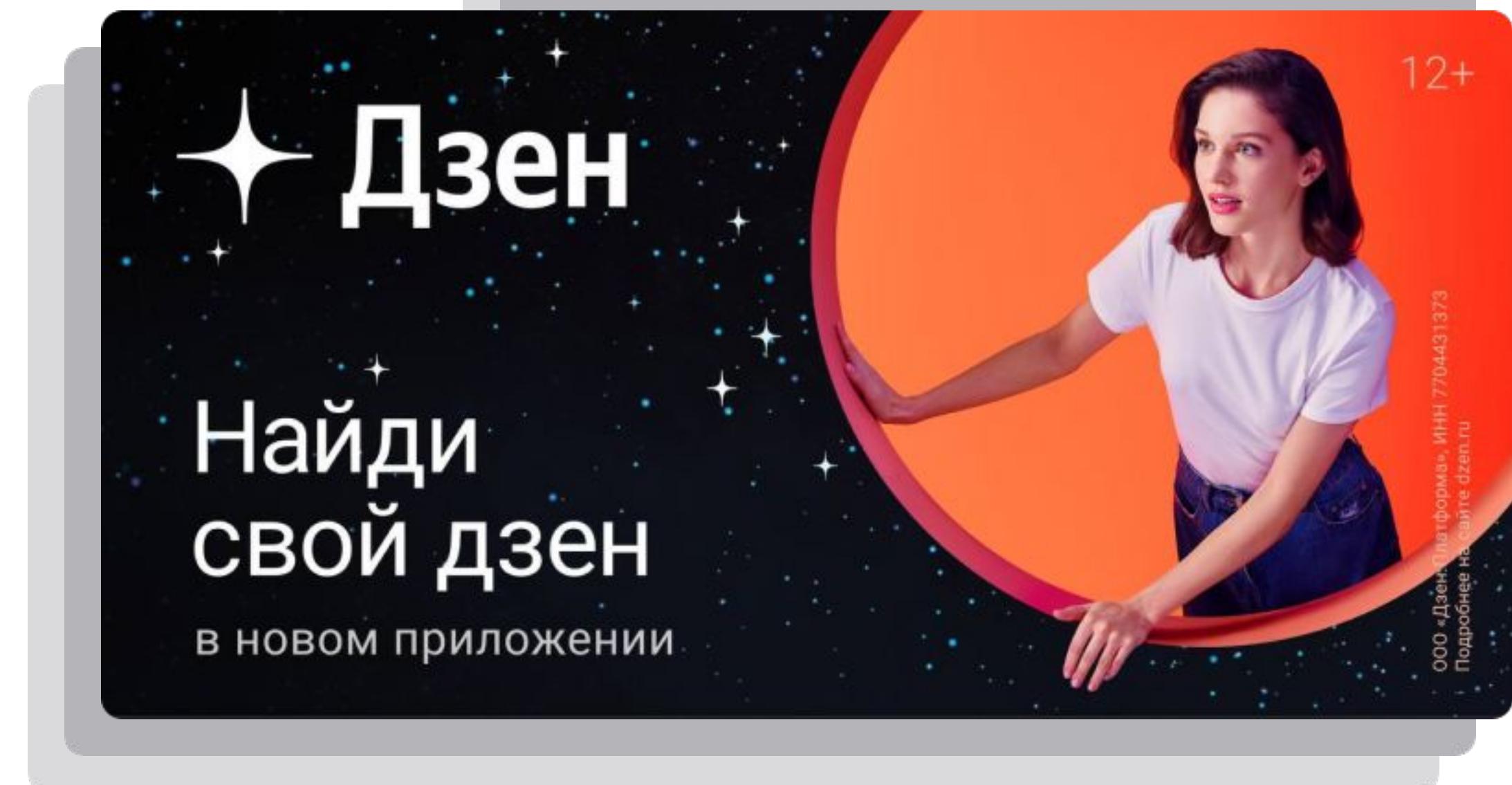
Saint
HighLoad++
2023



Top-level метрики сервиса

Как растить? DAU

- ❖ Маркетинговыми кампаниями
- ❖ Эксклюзивным контентом
- ❖ SEO



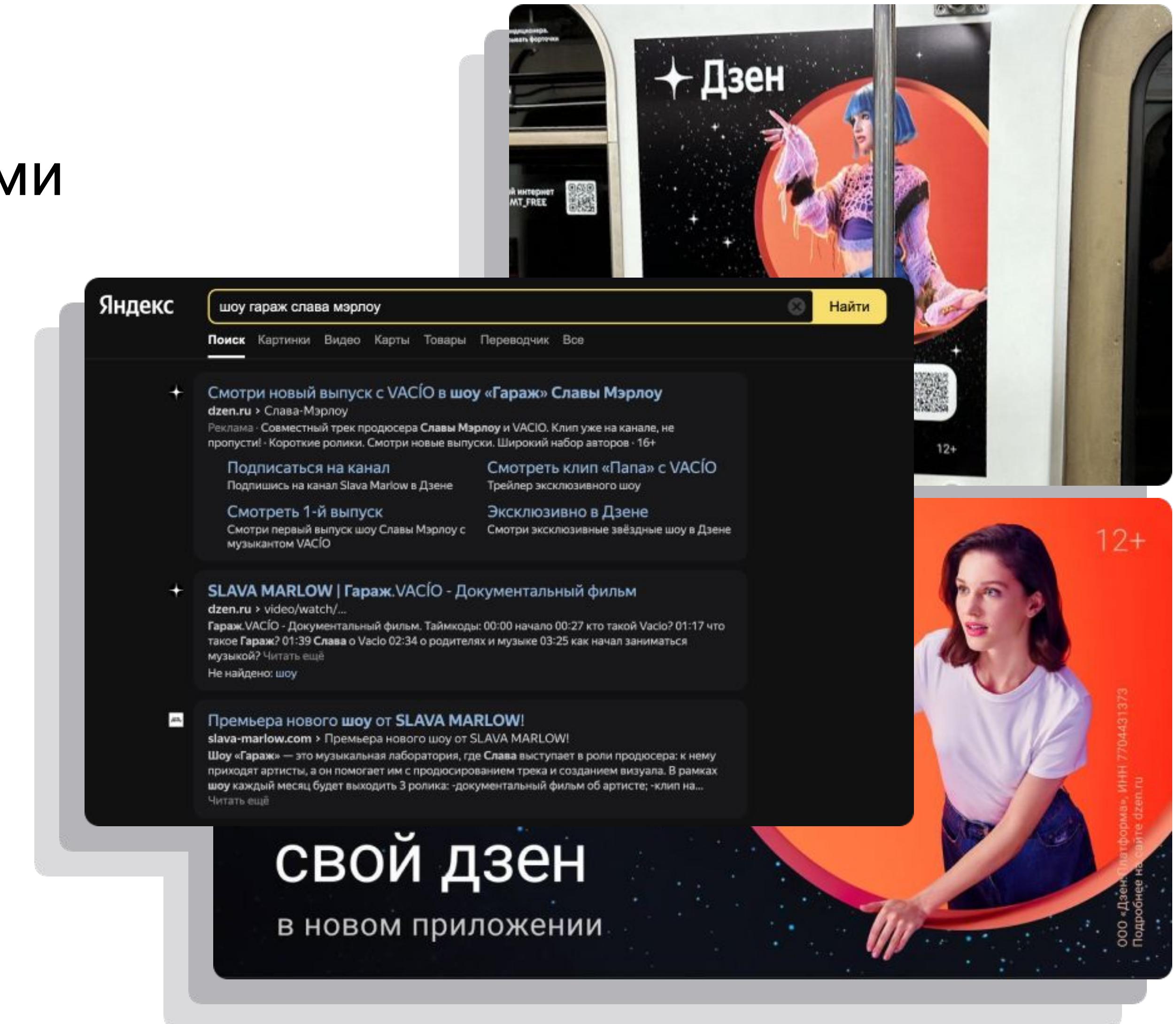
Saint
HighLoad++
2023



Top-level метрики сервиса

Как растить? DAU

- ❖ Маркетинговыми кампаниями
- ❖ Эксклюзивным контентом
- ❖ SEO



Дзен

Saint
HighLoad++
2023



Top-level метрики сервиса

Как растить? Timespent

- ★ Повышать качество предсказания
 - ★ Лучше интерпретировать сигнал
 - ★ Пробовать новые сложные модели
- ★ Увеличивать контентную базу



Saint
HighLoad++
2023



Top-level метрики сервиса

Как растить? NPS

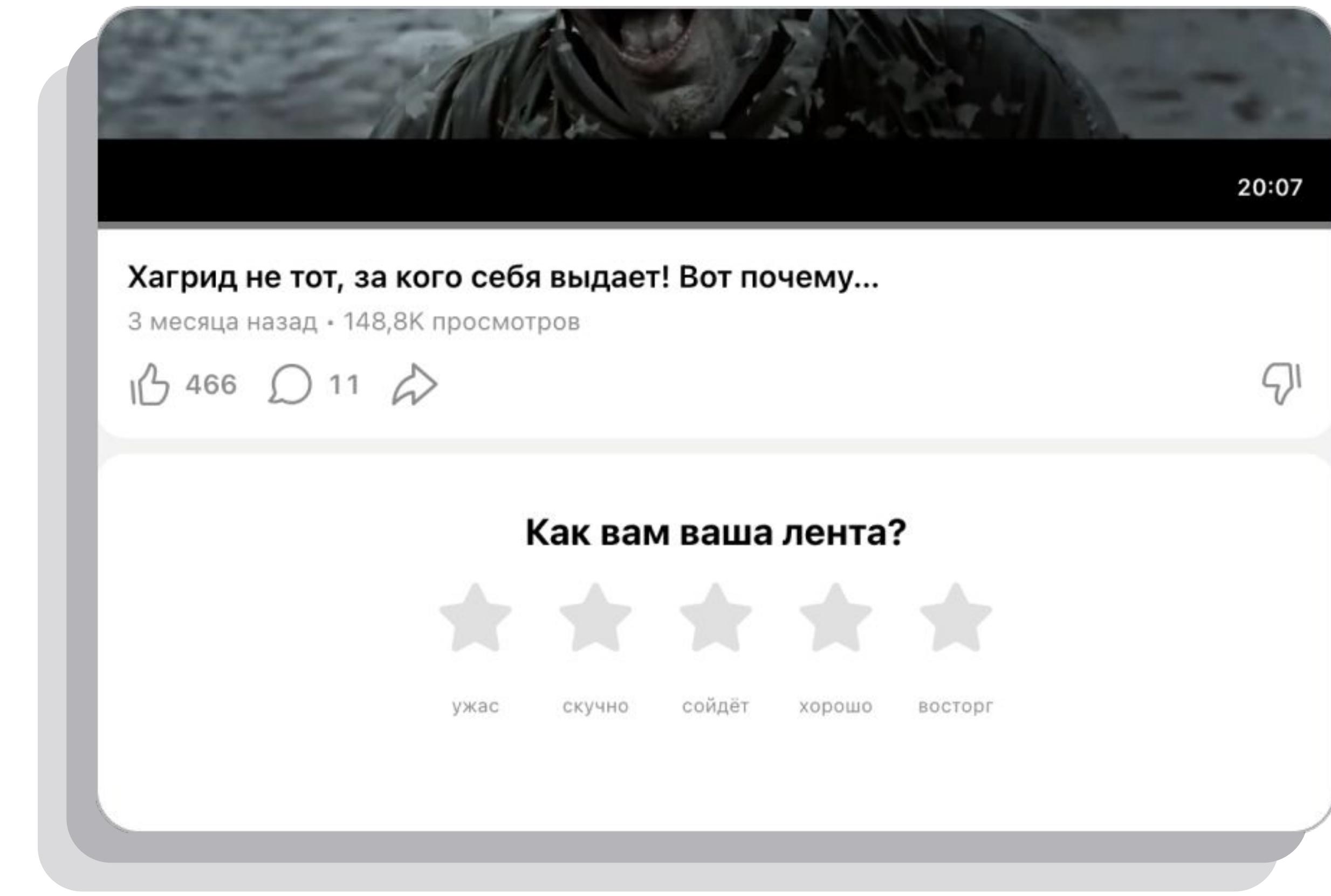
Индекс NPS (англ. Net Promoter Score) – индекс определения приверженности потребителей товару или компании / бренду (индекс готовности рекомендовать)



Top-level метрики сервиса

Как растить? Net Promoter Score

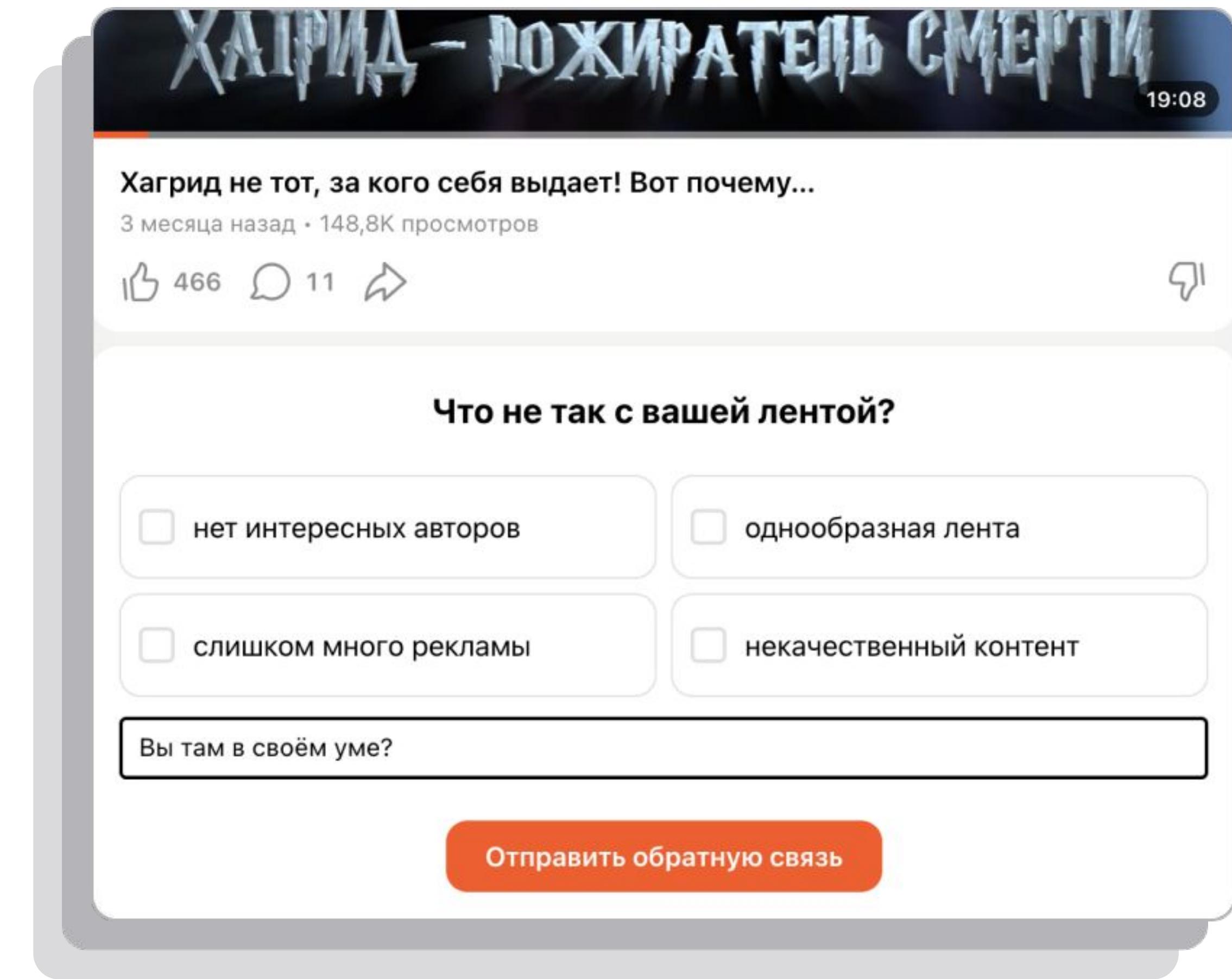
- ★ Провести исследование:
спросить пользователей



Top-level метрики сервиса

Как растить? Net Promoter Score

- ★ Провести исследование:
спросить пользователей



Top-level метрики сервиса

NPS, результаты исследования:

- ★ Недовольство частью ленты:
 - ★ «Не понимаю, почему это показываете, я такое не смотрел»
- ★ Не все интересы пользователя есть в ленте
 - ★ «Я иногда читаю автоспорт, но в ленте его почти нет»



Saint
HighLoad++
2023



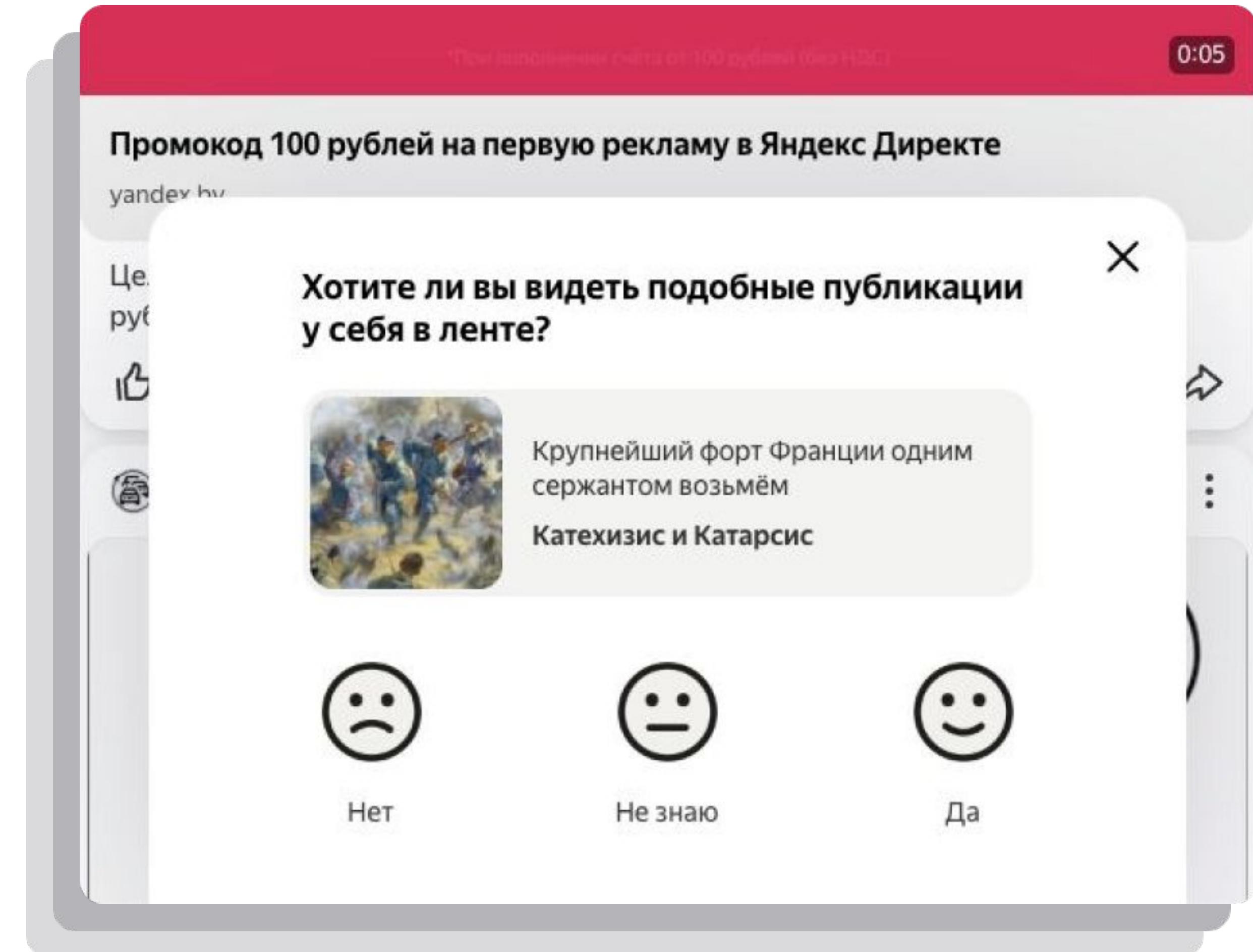
Top-level метрики сервиса

NPS, результаты исследования:

- ★ Недовольство частью ленты:
 - ★ «Не понимаю, почему это показываете, я такое не смотрел»
- ★ Не все интересы пользователя есть в ленте
 - ★ «Я иногда читаю автоспорт, но в ленте его почти нет»
- ★ Рекламное давление
- ★ Баги
- ★ Погода
- ★ etc

Что хотим?

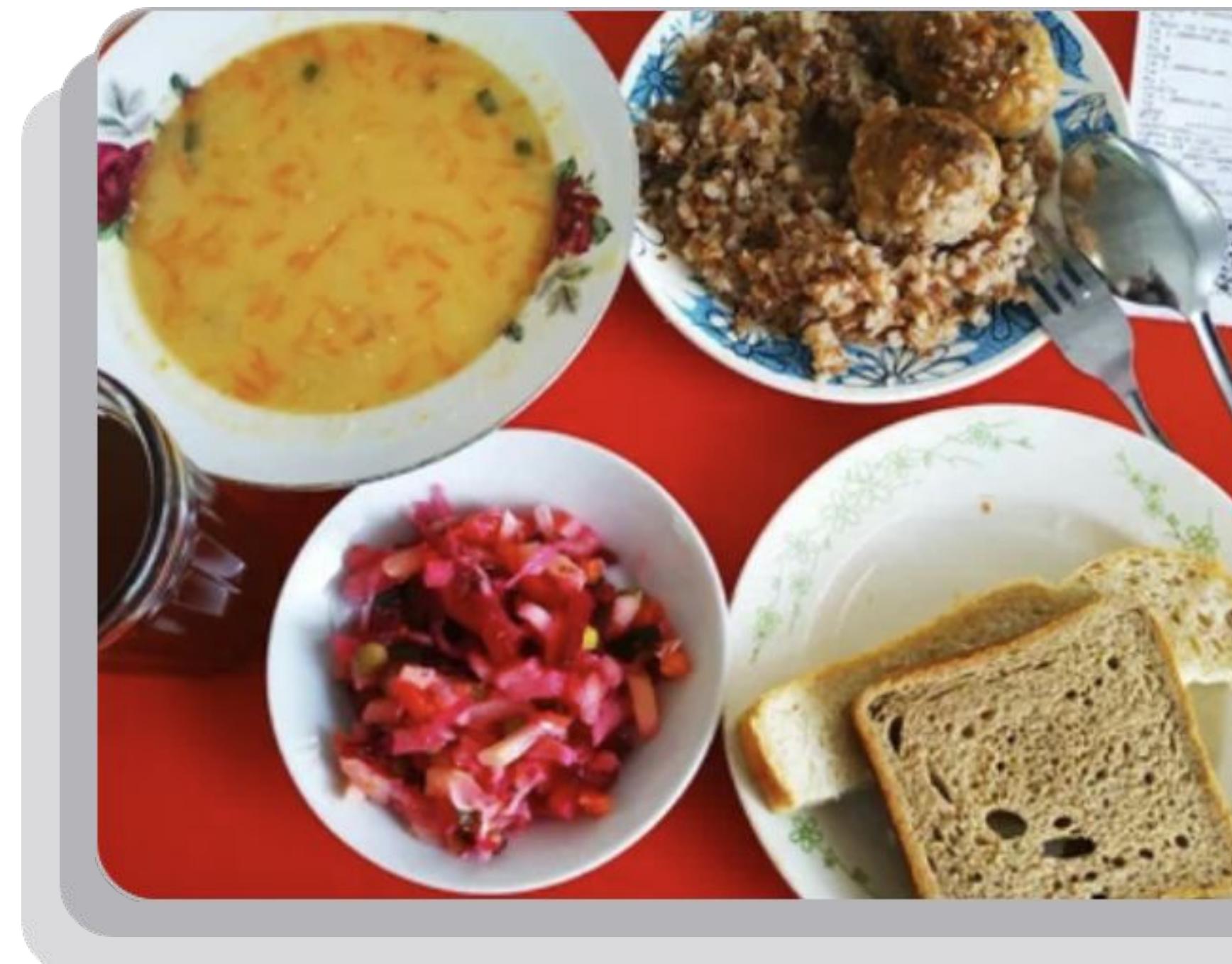
Вырастить
чувствительную прокси-
метрику к NPS – CSI-
рекомендаций



Давайте подебажим

«Не понимаю, почему это
показываете, я такое не смотрел»

(с) финансовый директор



Скромный дег... [Подписаться](#) ...
10K subscribers

Обед за 176 рублей. Подорожал
за год всего на рубль. Столовка
из СССР. Неужели мяса стали...

Сяду на пенек, съем пирожок Ну, про
год это я "завернул" конечно. Чуть...

2 months ago



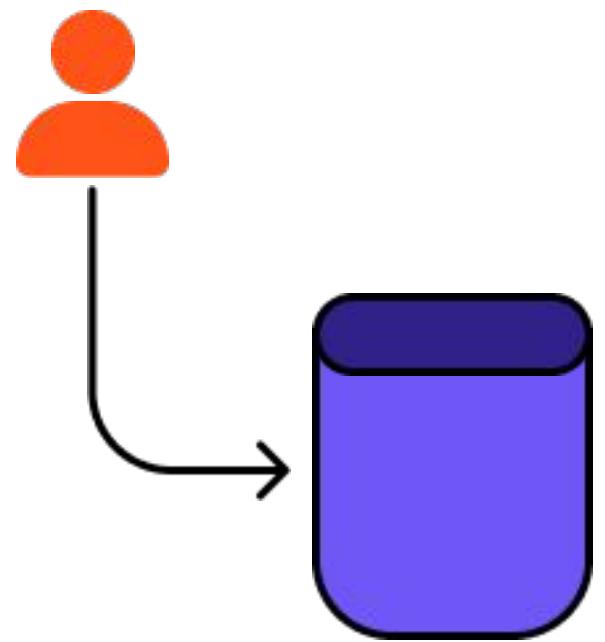
Дзен

Saint
HighLoad++
2023



Схема устройства рекомендаций

Сбор
и хранение
историй



Saint
HighLoad++
2023



Схема устройства рекомендаций

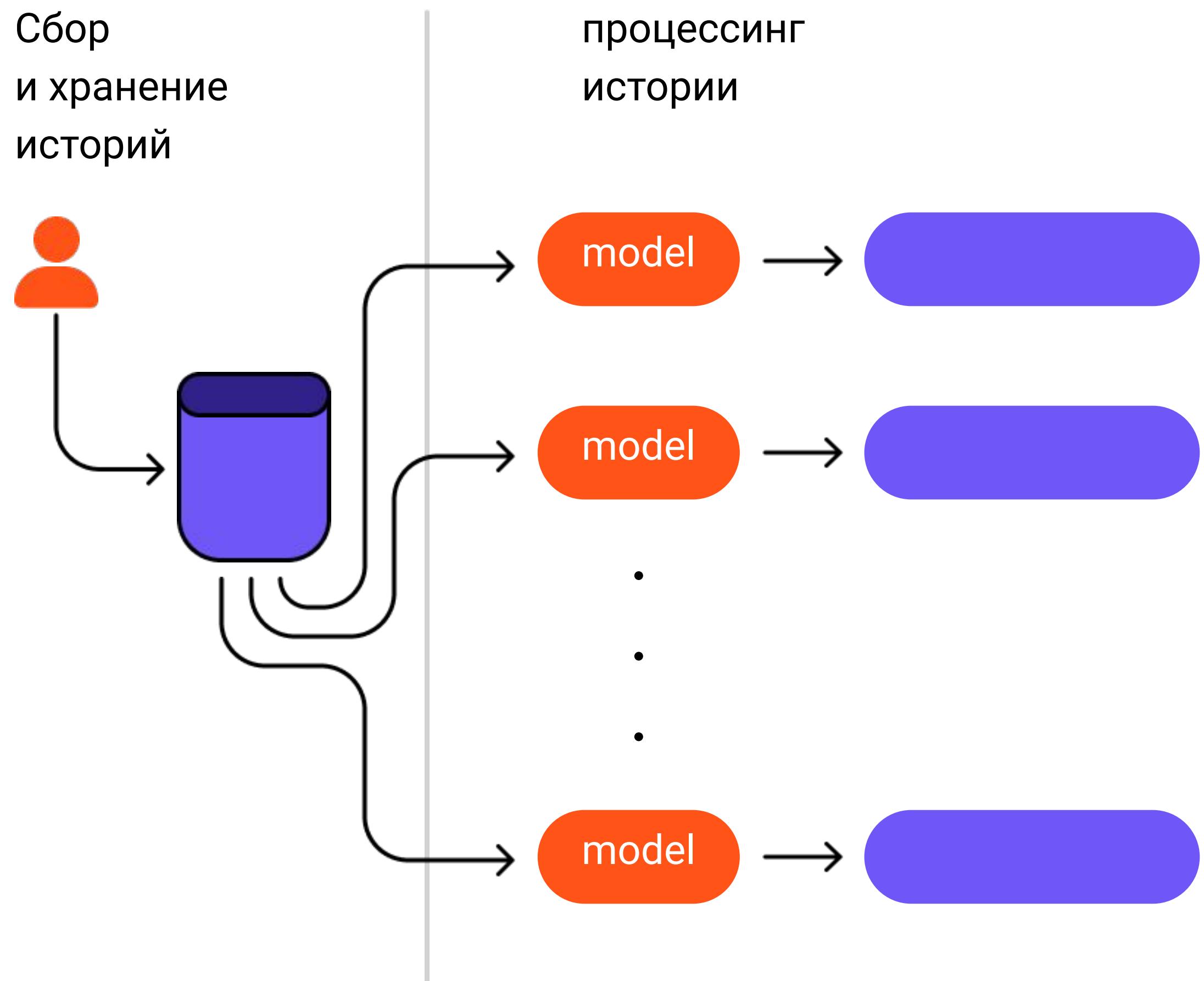


Схема устройства рекомендаций

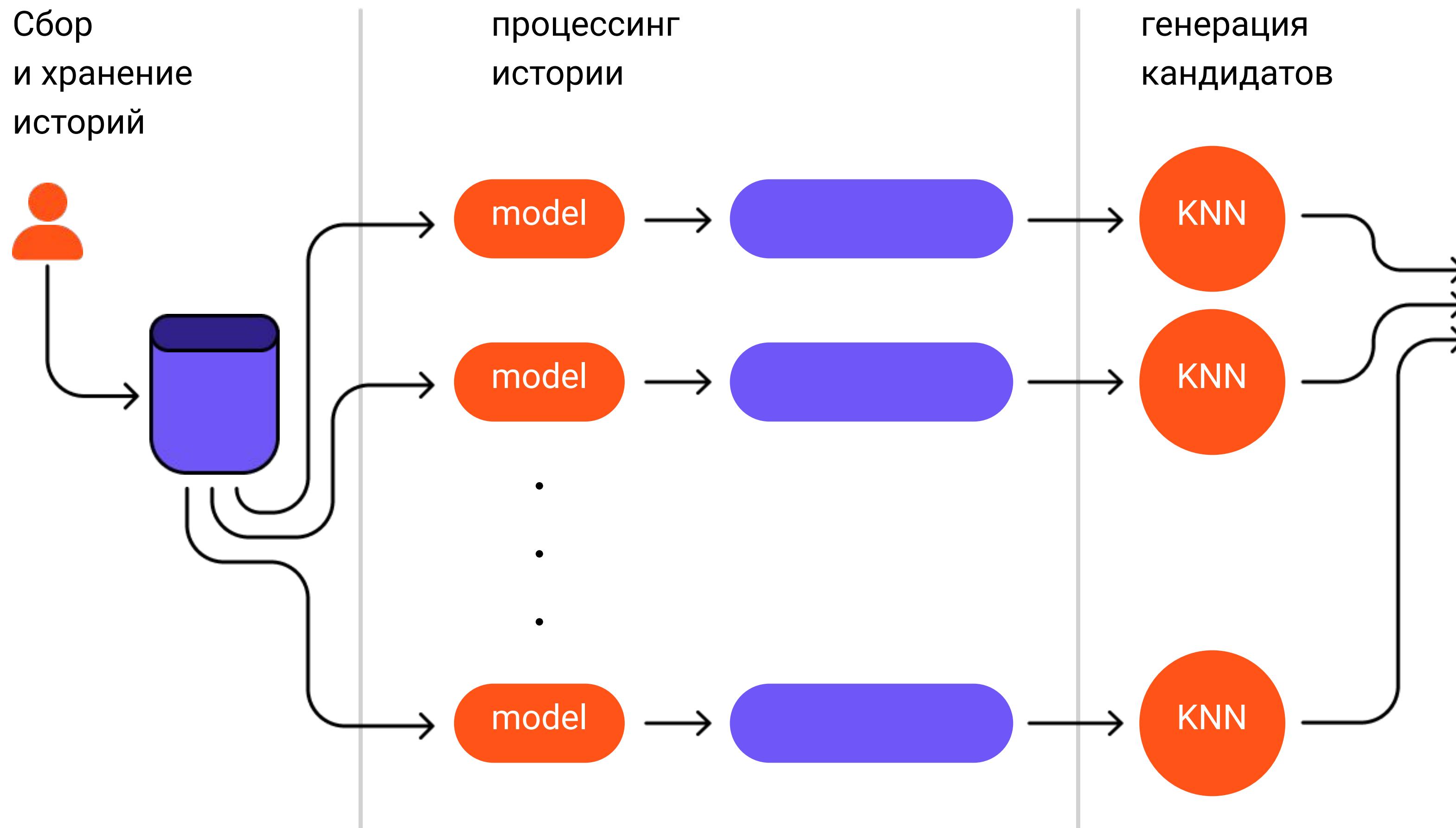


Схема устройства рекомендаций

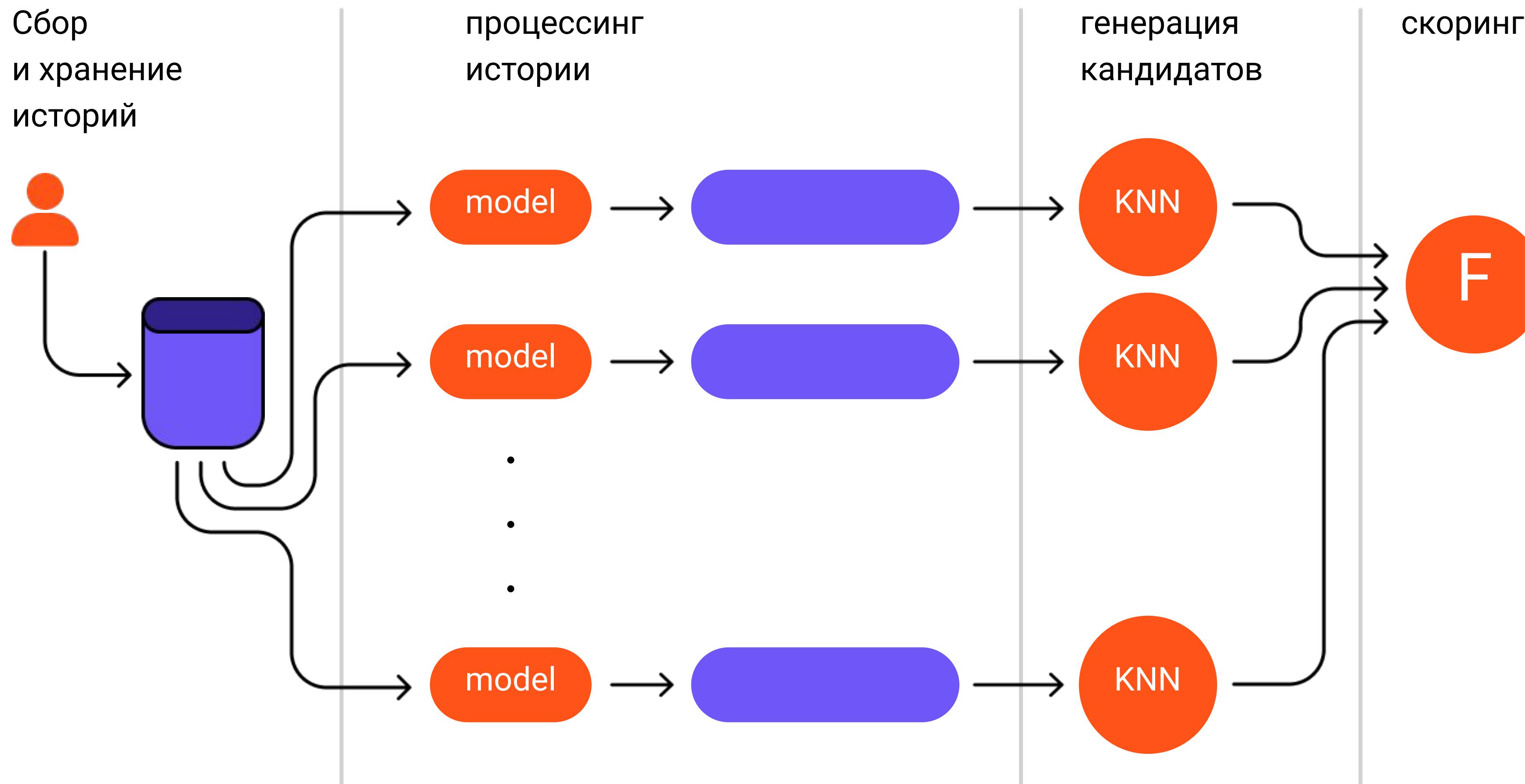


Схема устройства рекомендаций

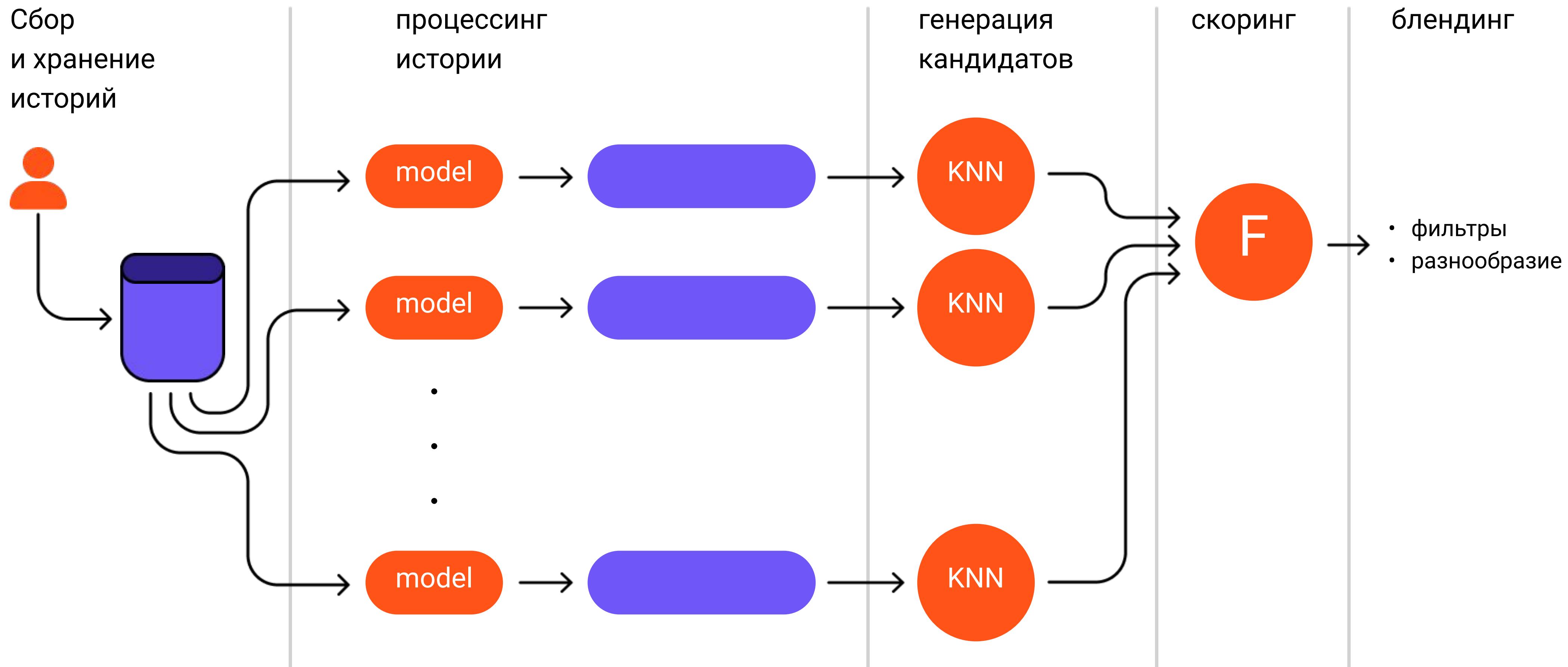


Схема устройства рекомендаций

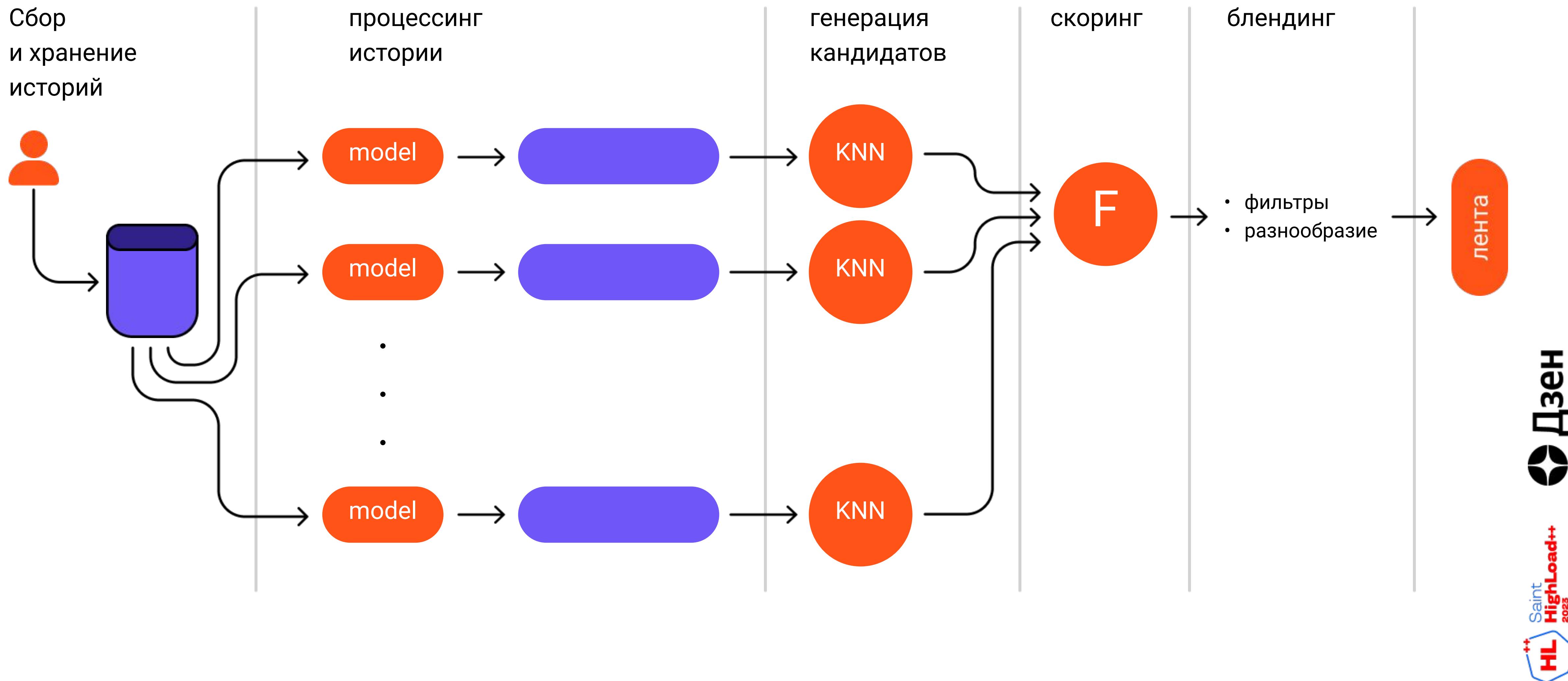
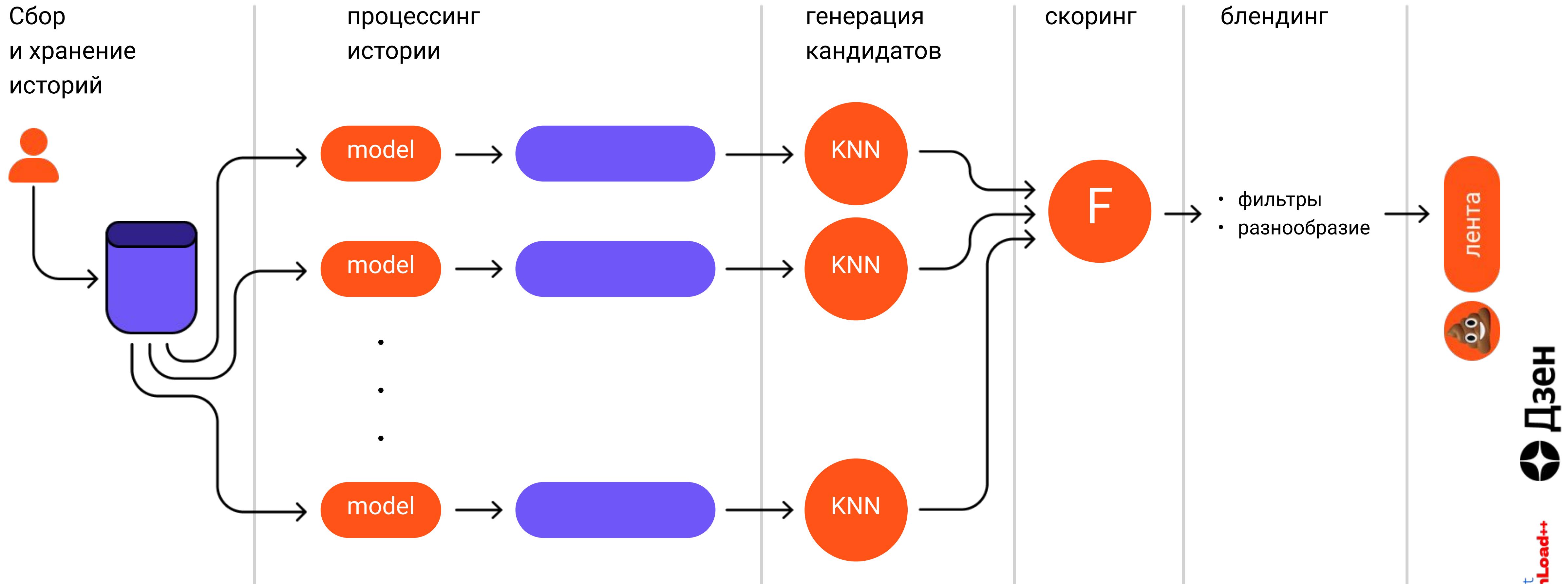
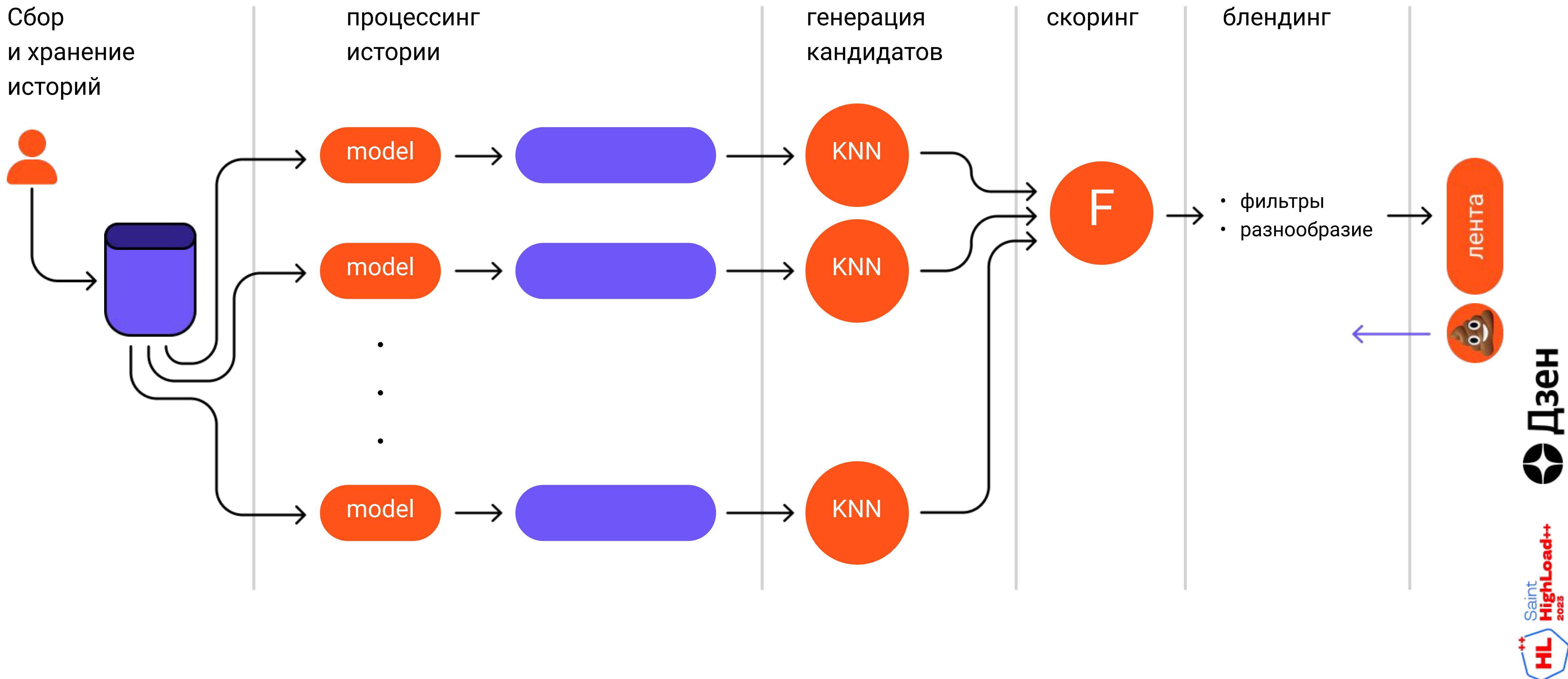


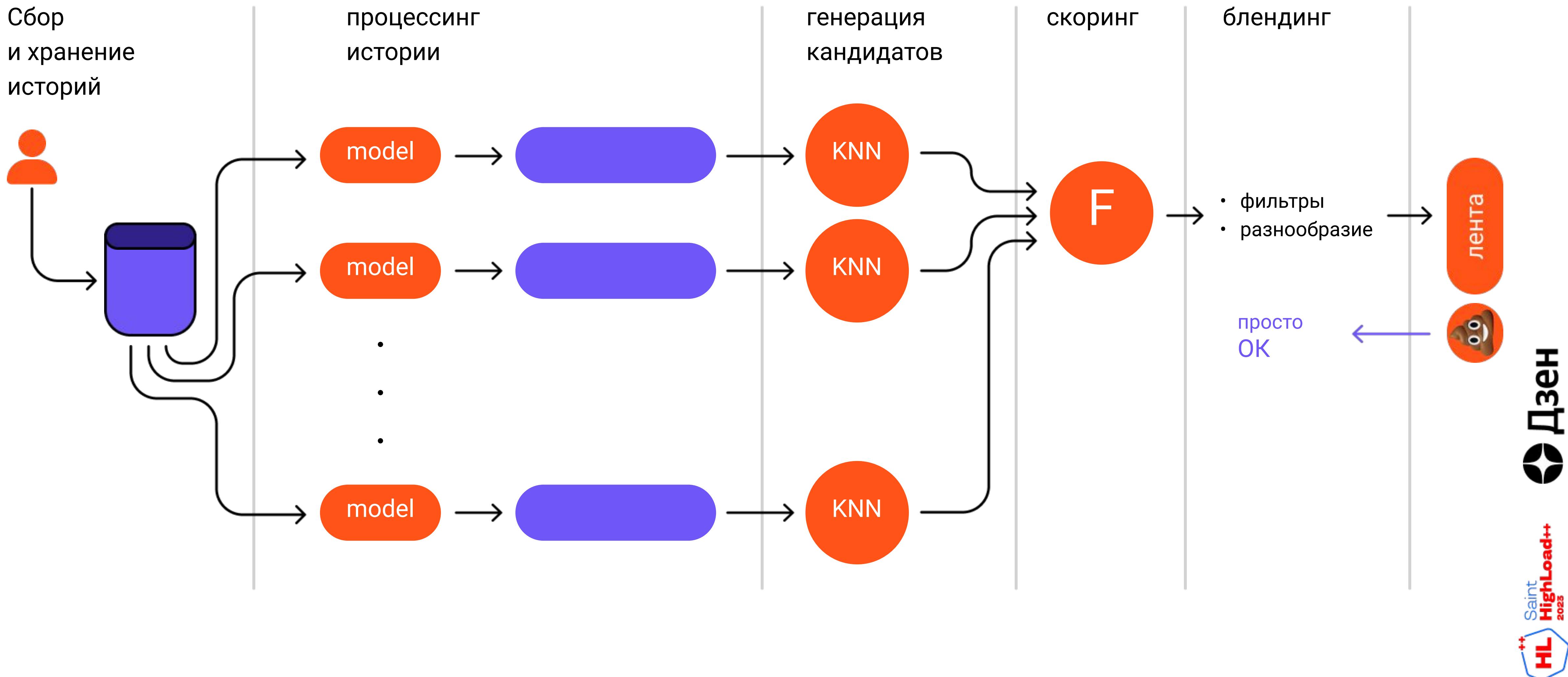
Схема устройства рекомендаций



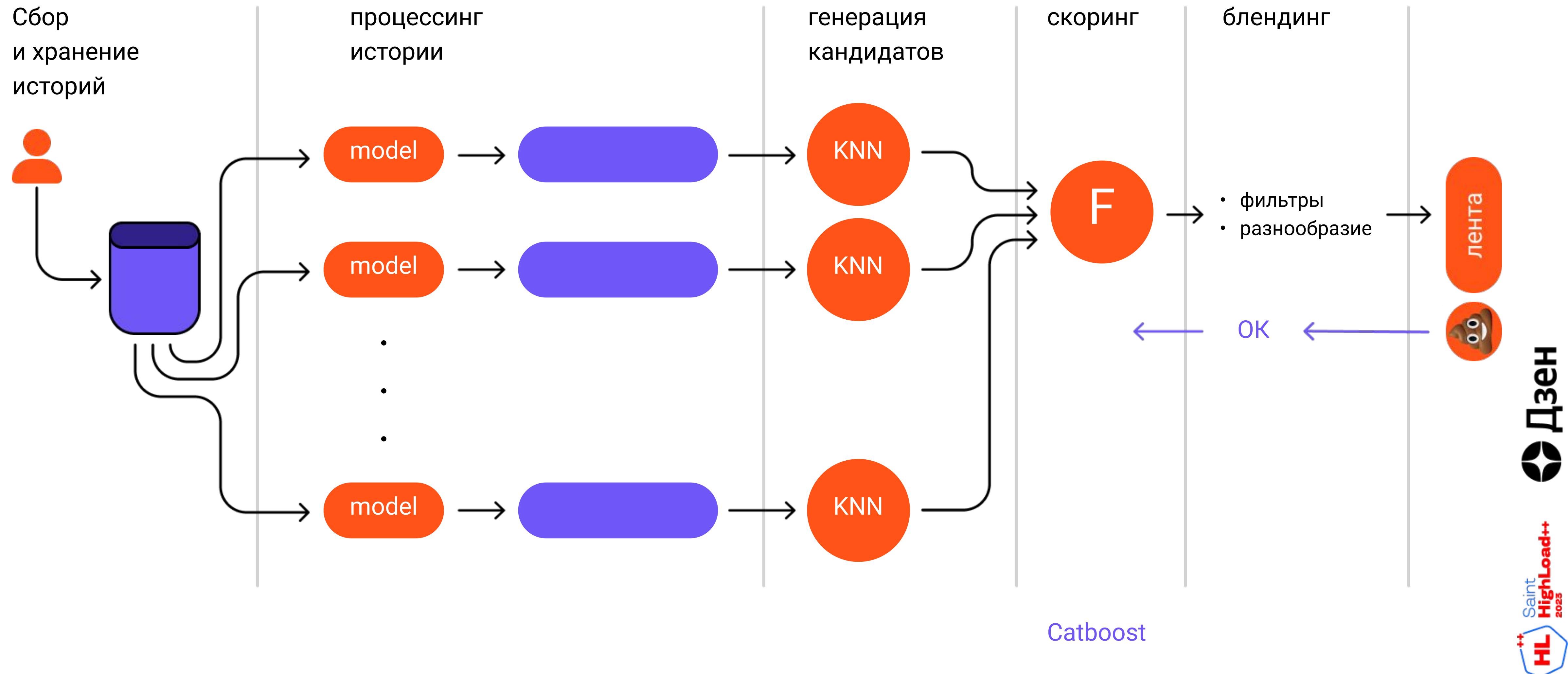
Дебаг ленты



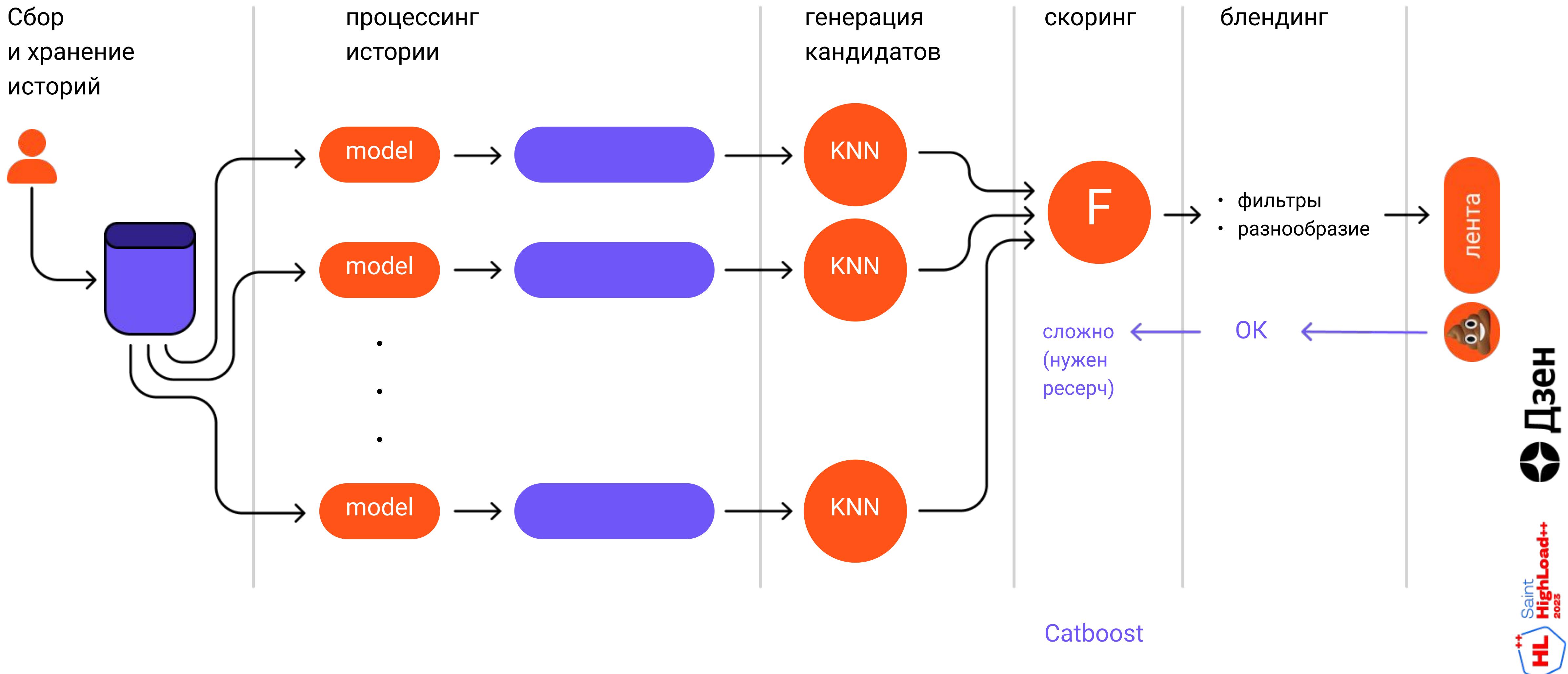
Дебаг ленты



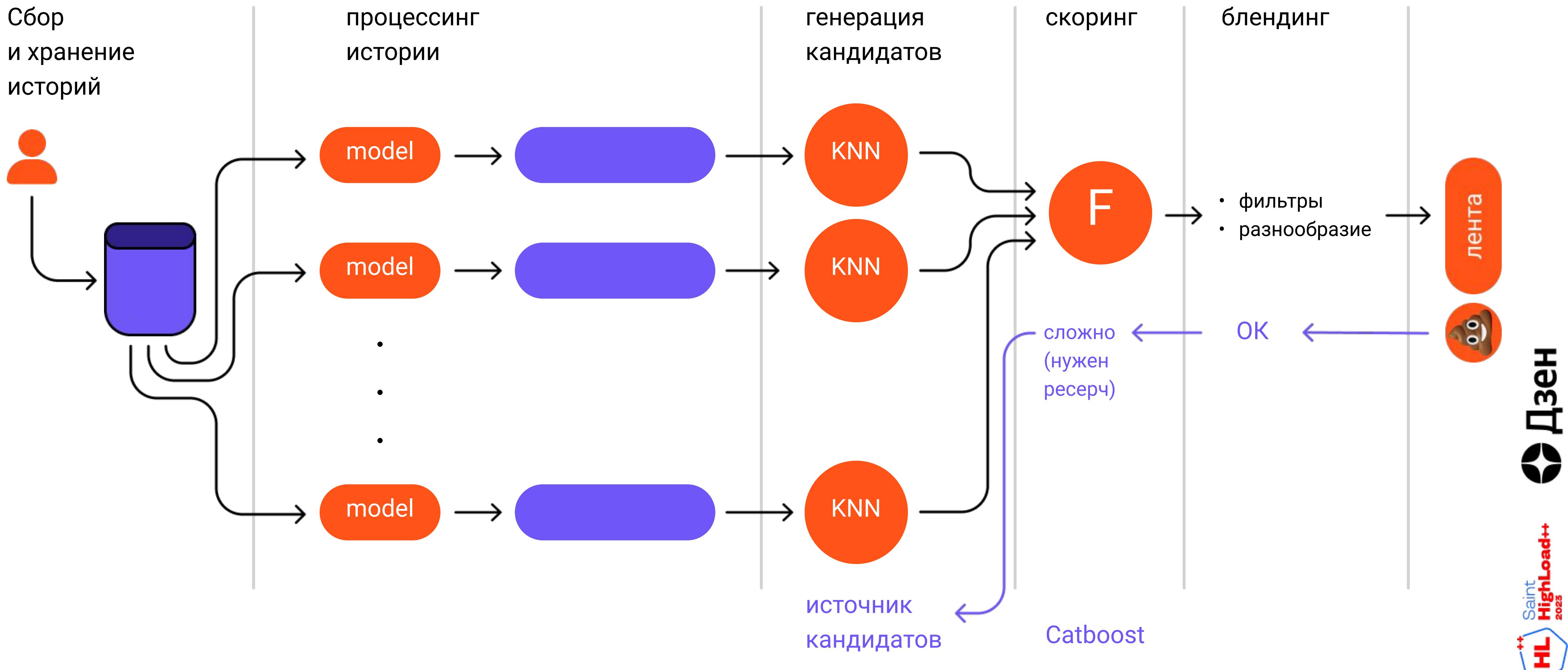
Дебаг ленты



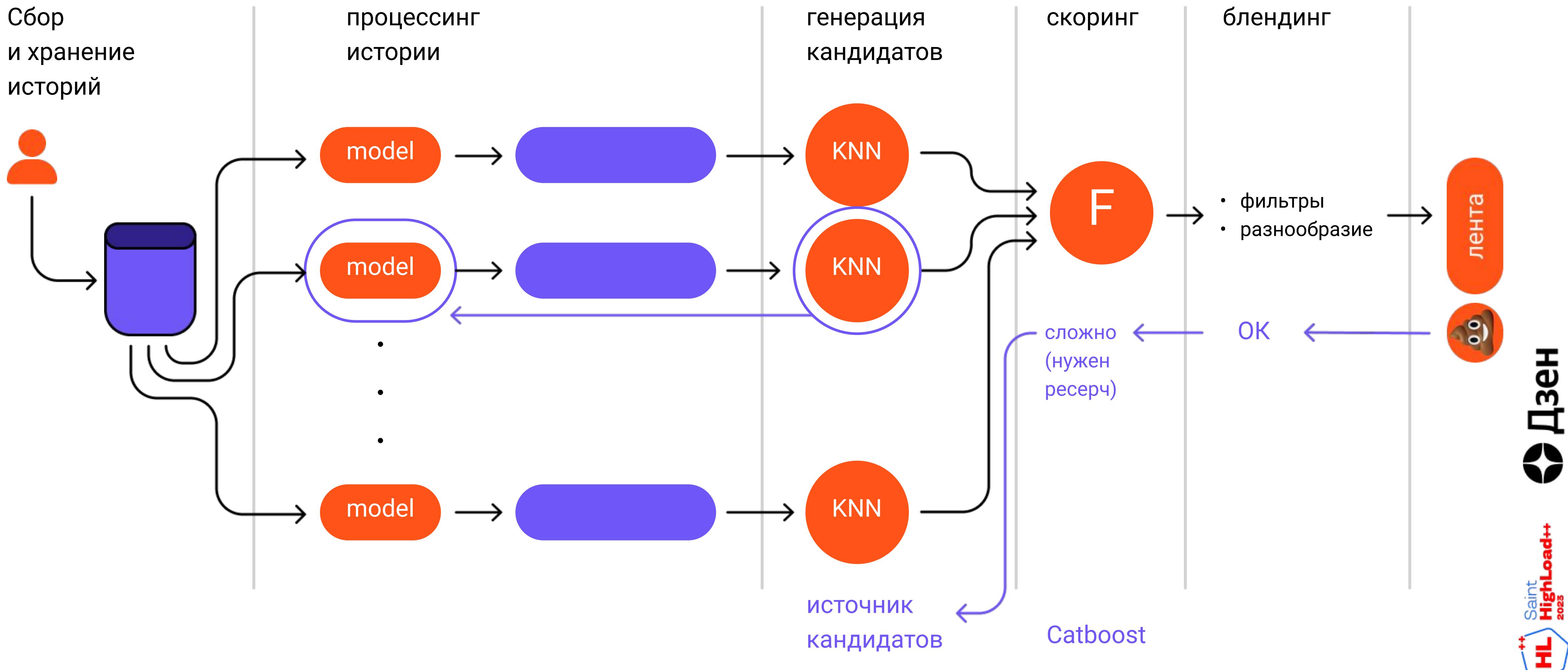
Дебаг ленты



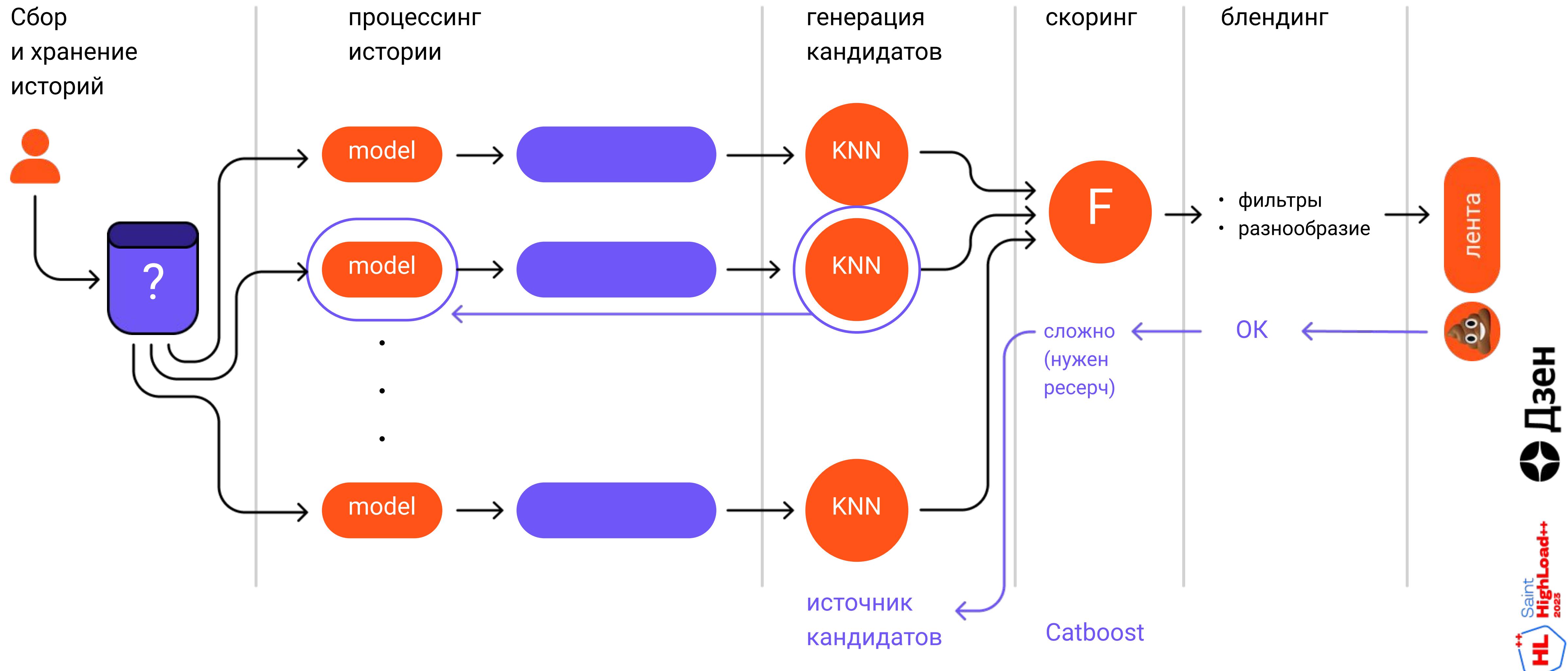
Дебаг ленты



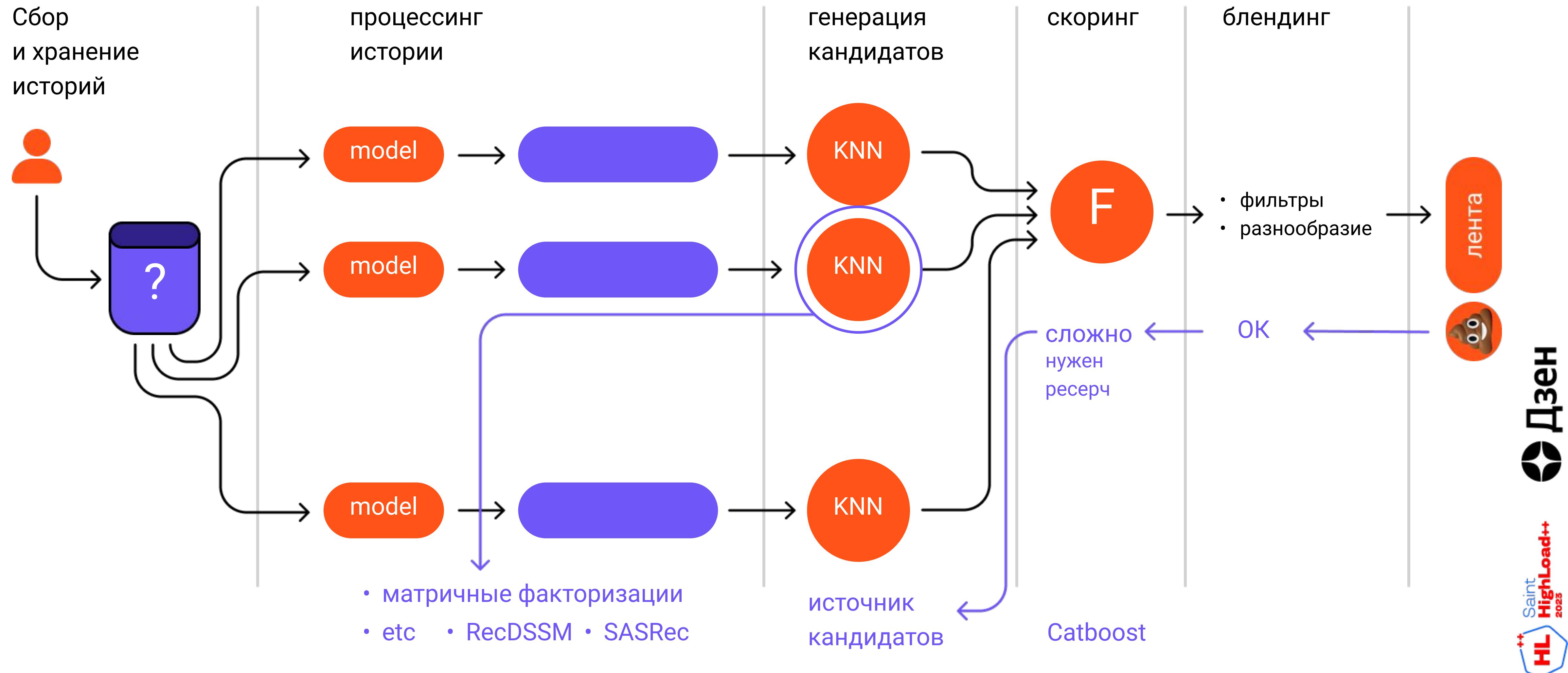
Дебаг ленты



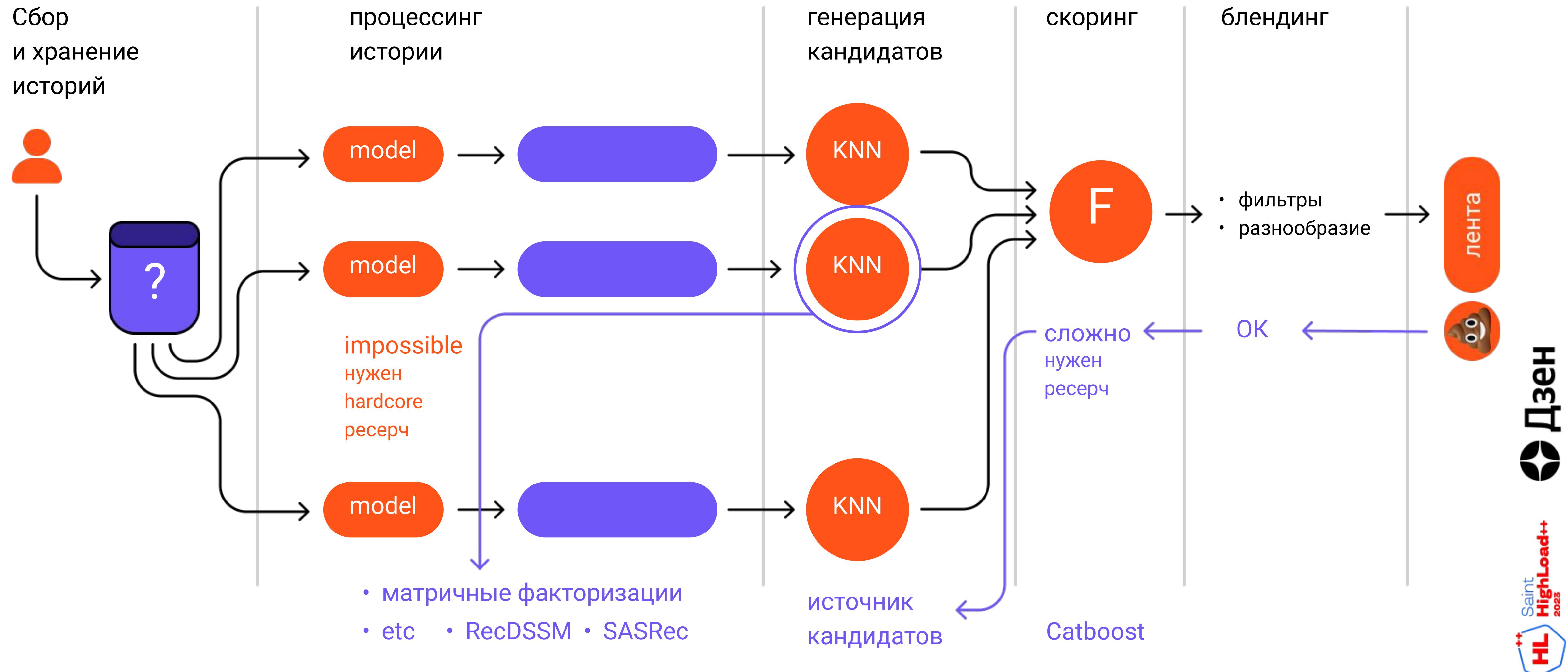
Дебаг ленты



Дебаг ленты



Дебаг ленты



Что хотим?

Сделать этап генерации
кандидатов **простым**



Saint
HighLoad++
2023



Проблемы

Часть ленты совсем не похожа
на историю пользователя

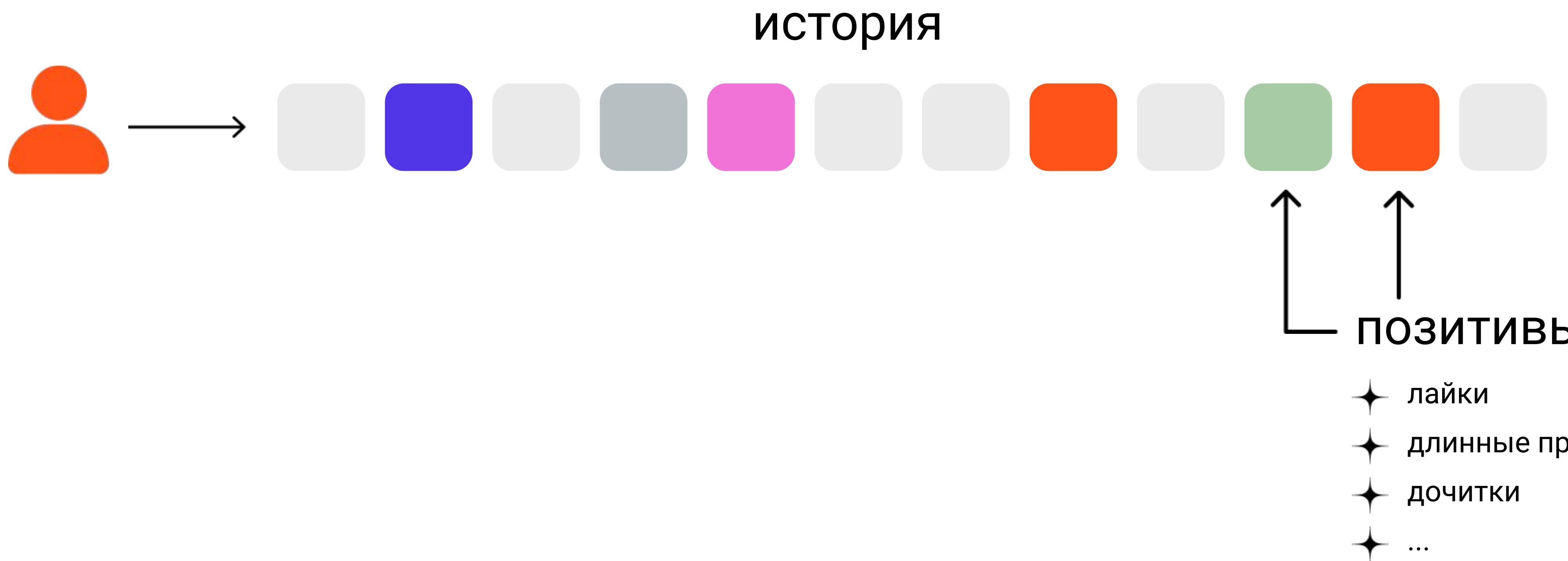


Saint
HighLoad++
2023



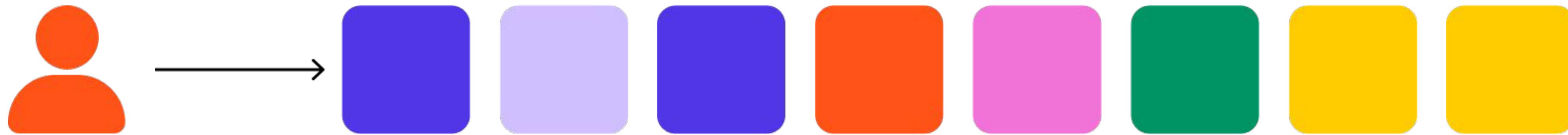
Проблемы

Часть ленты совсем не похожа
на историю пользователя

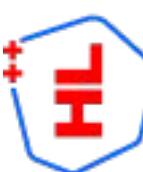


Проблемы

Часть ленты совсем не похожа
на историю пользователя

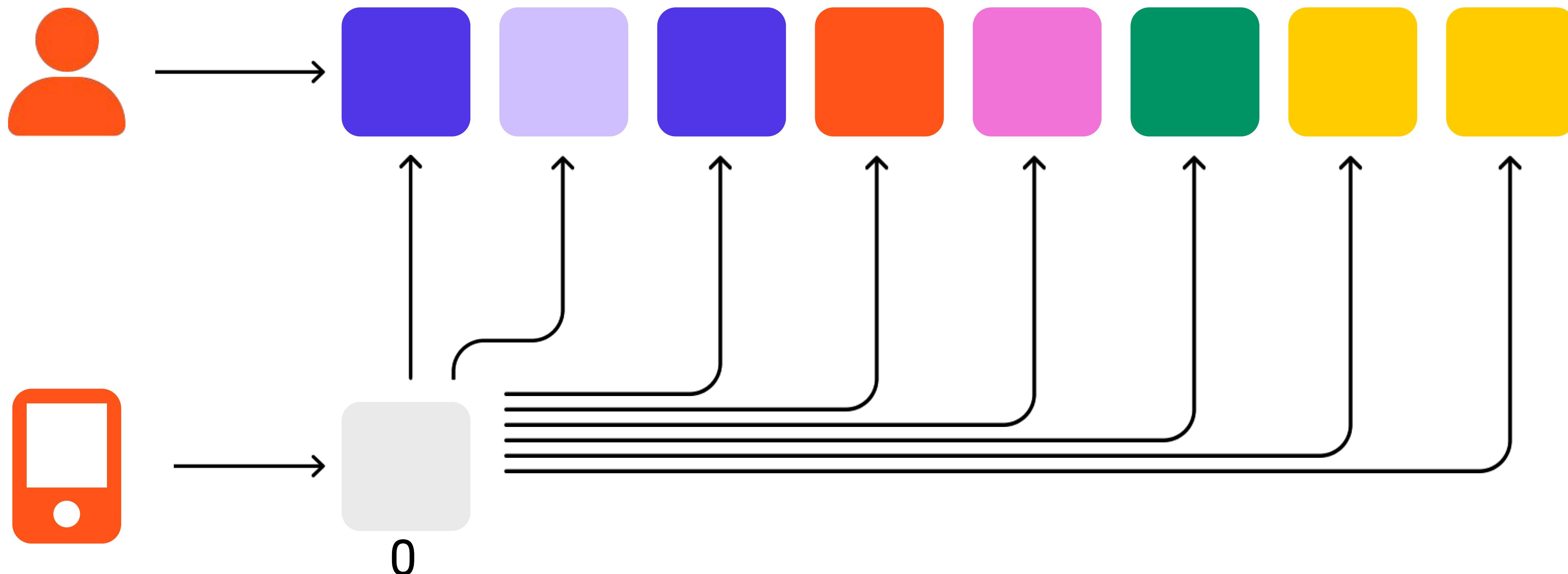


Saint
HighLoad++
2023



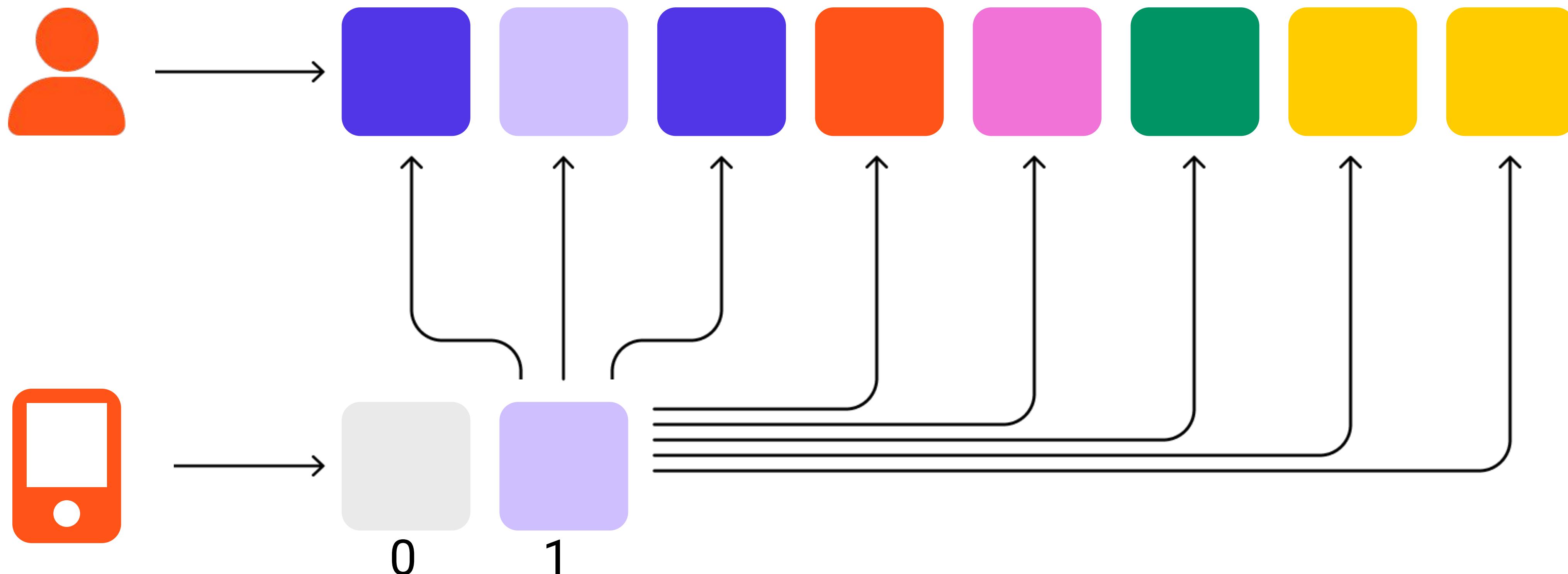
Проблемы

Часть ленты совсем не похожа
на историю пользователя



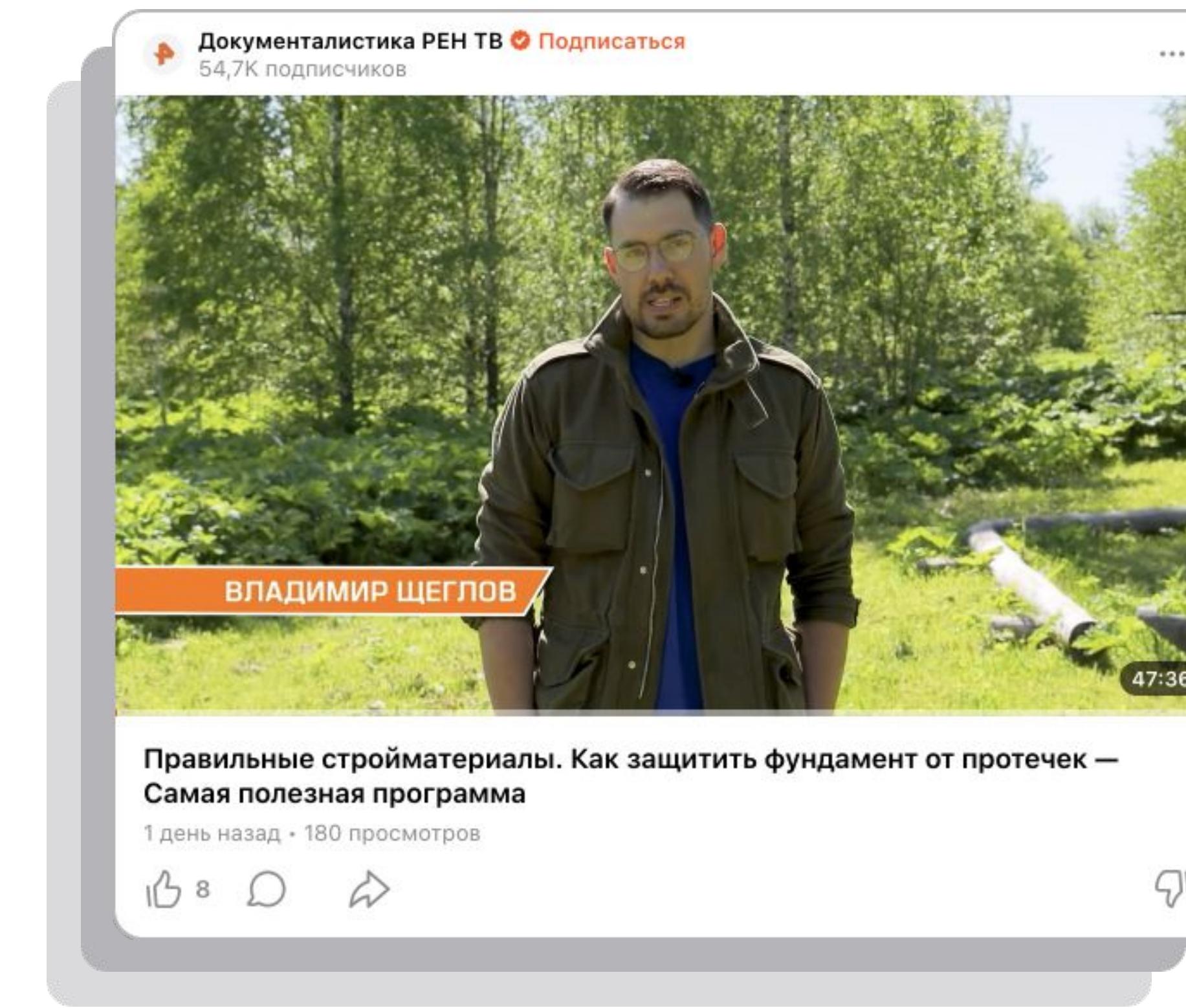
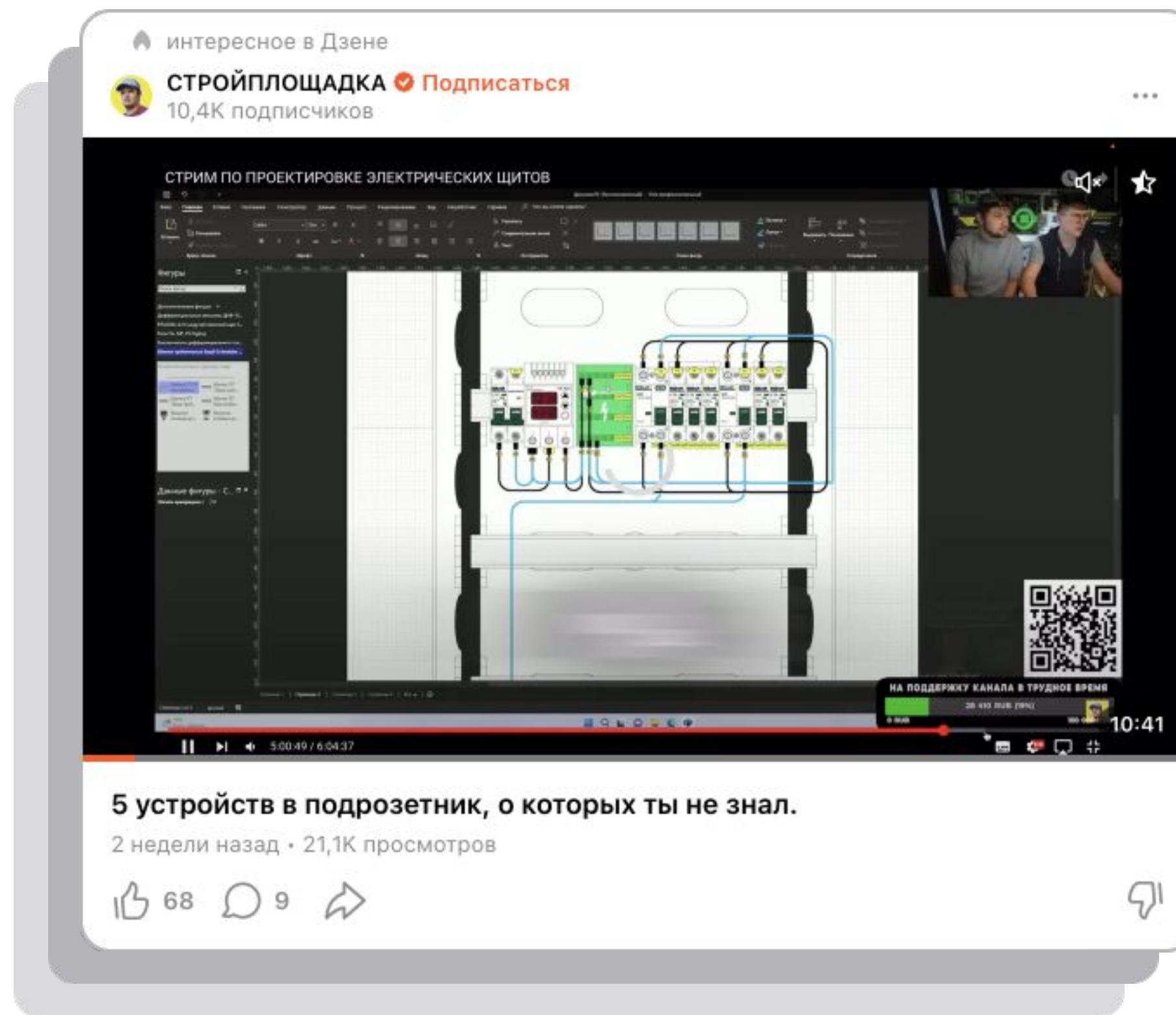
Проблемы

Часть ленты совсем не похожа
на историю пользователя



Как понять, что два айтема похожи?

Разметка похожестей краудом



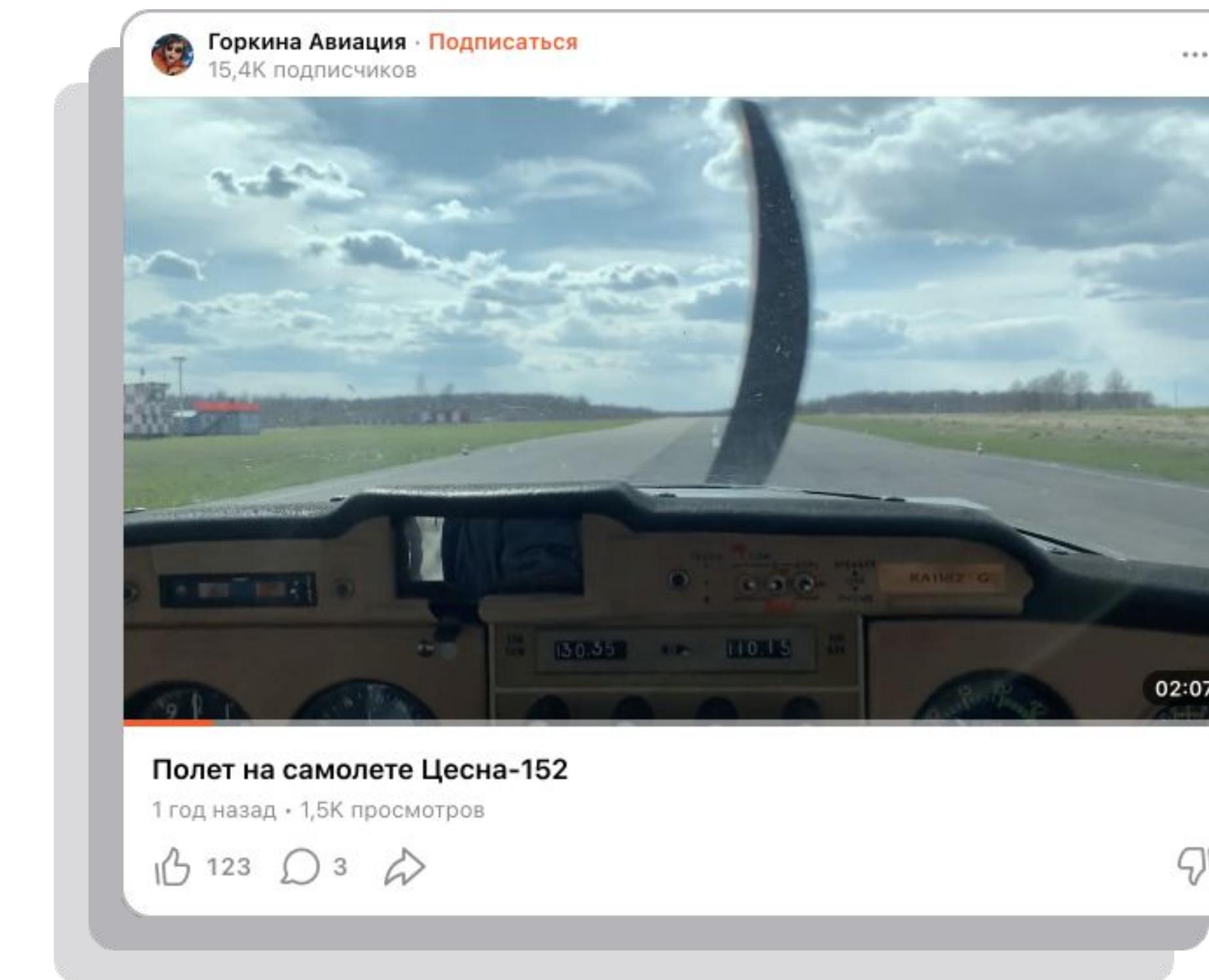
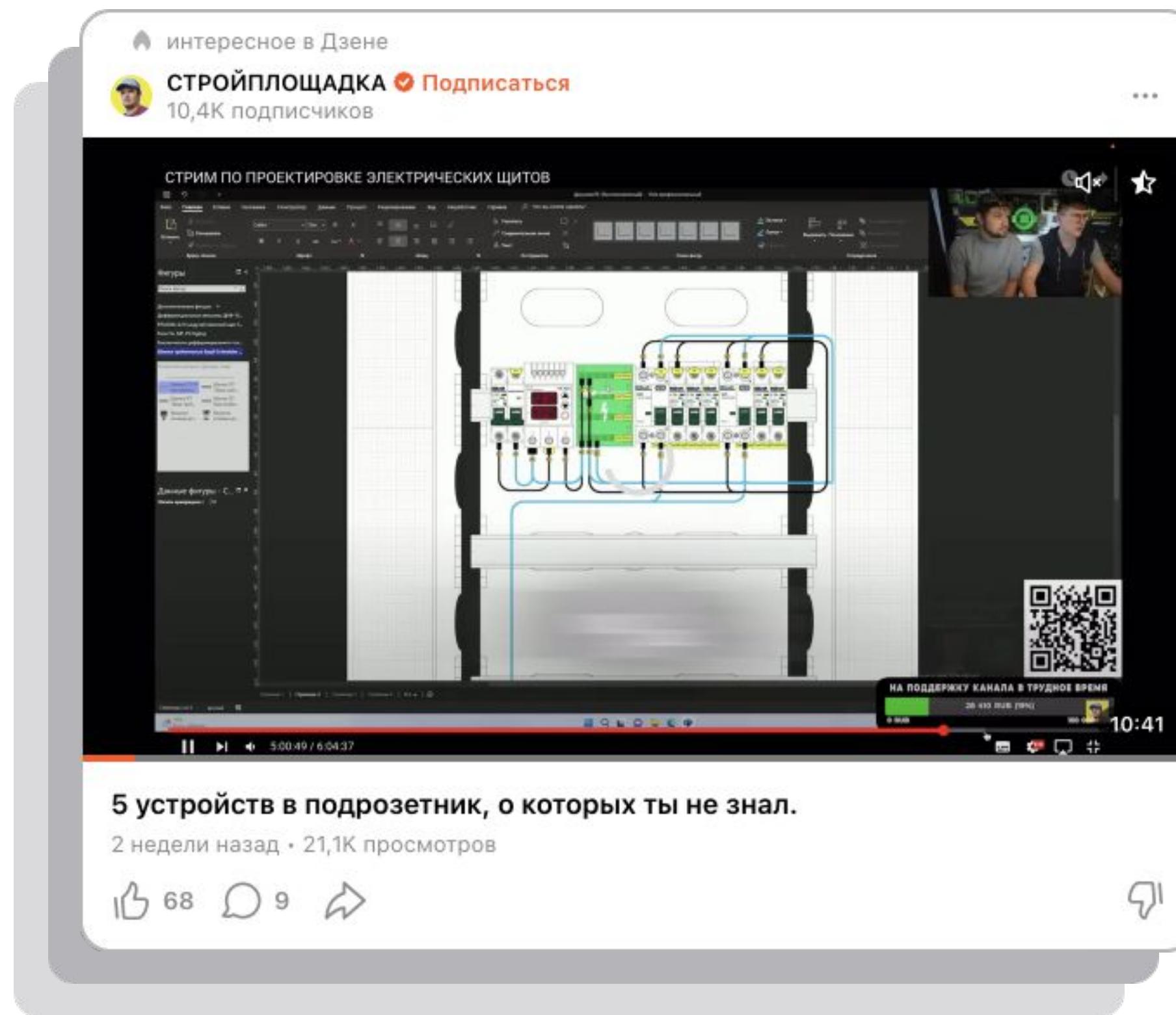
Дзен

Saint HighLoad++
2023



Как понять, что два айтема похожи?

Разметка похожестей краудом



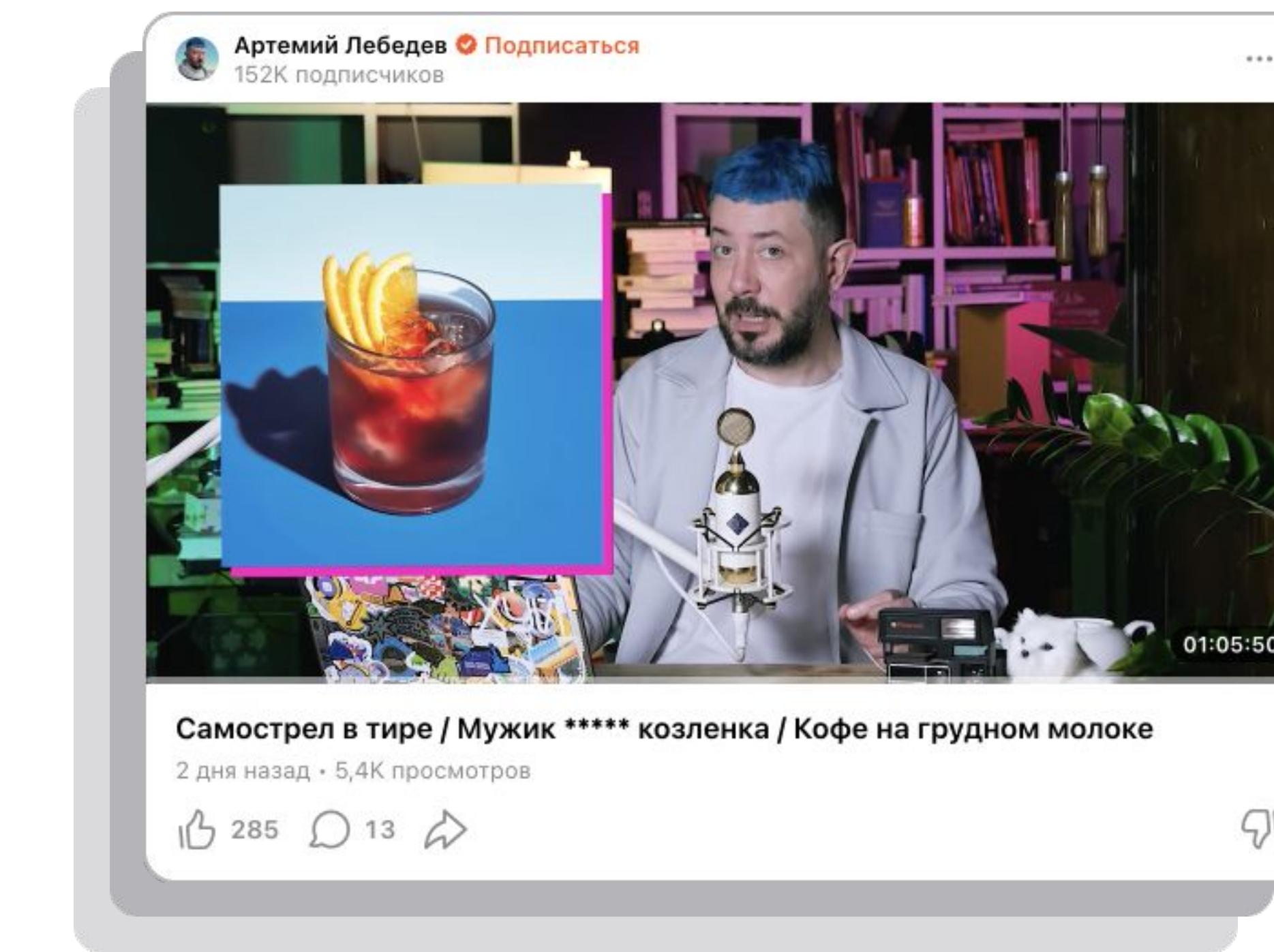
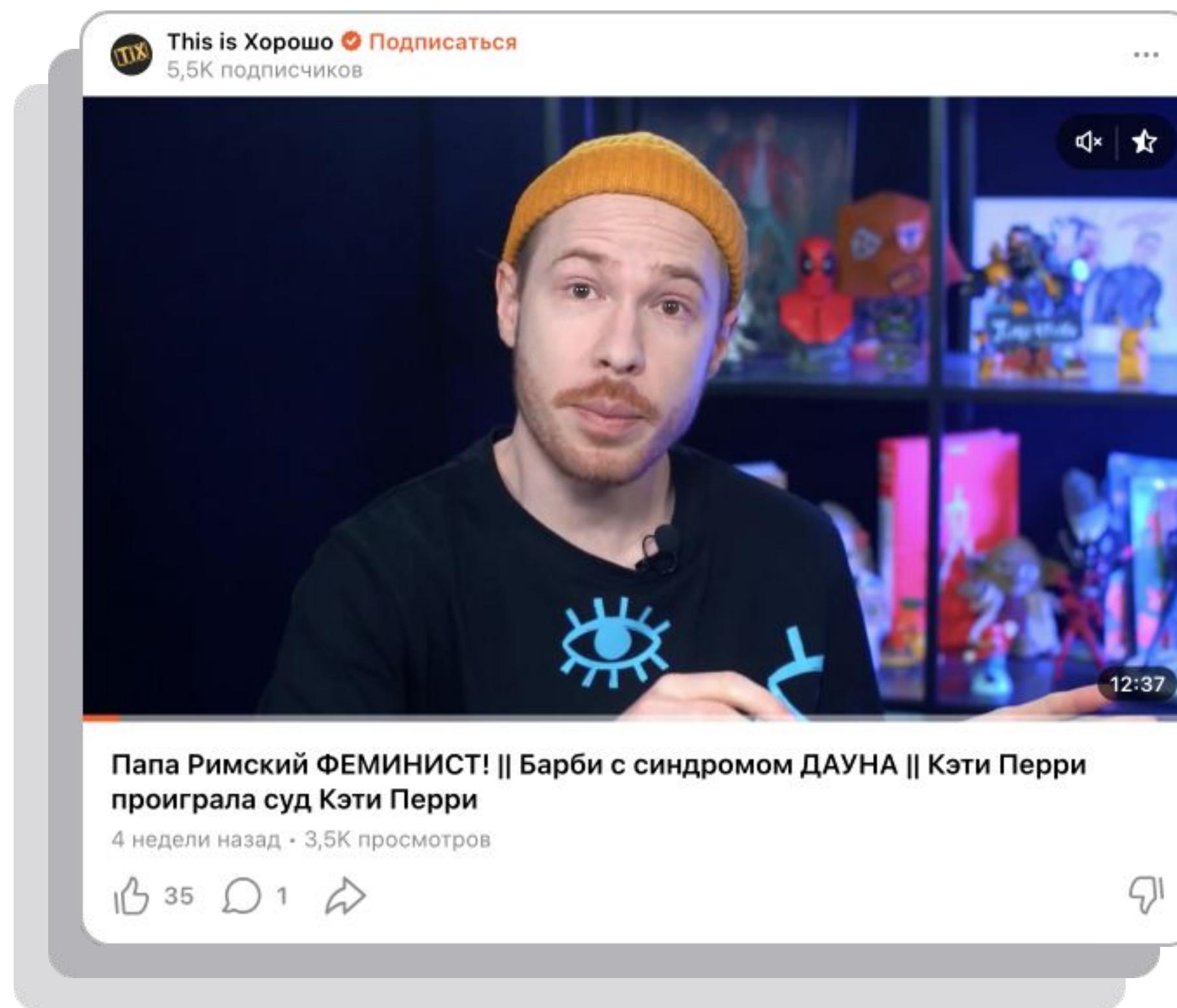
Дзен

Saint HighLoad++
2023



Как понять, что два айтема похожи?

Разметка похожестей краудом

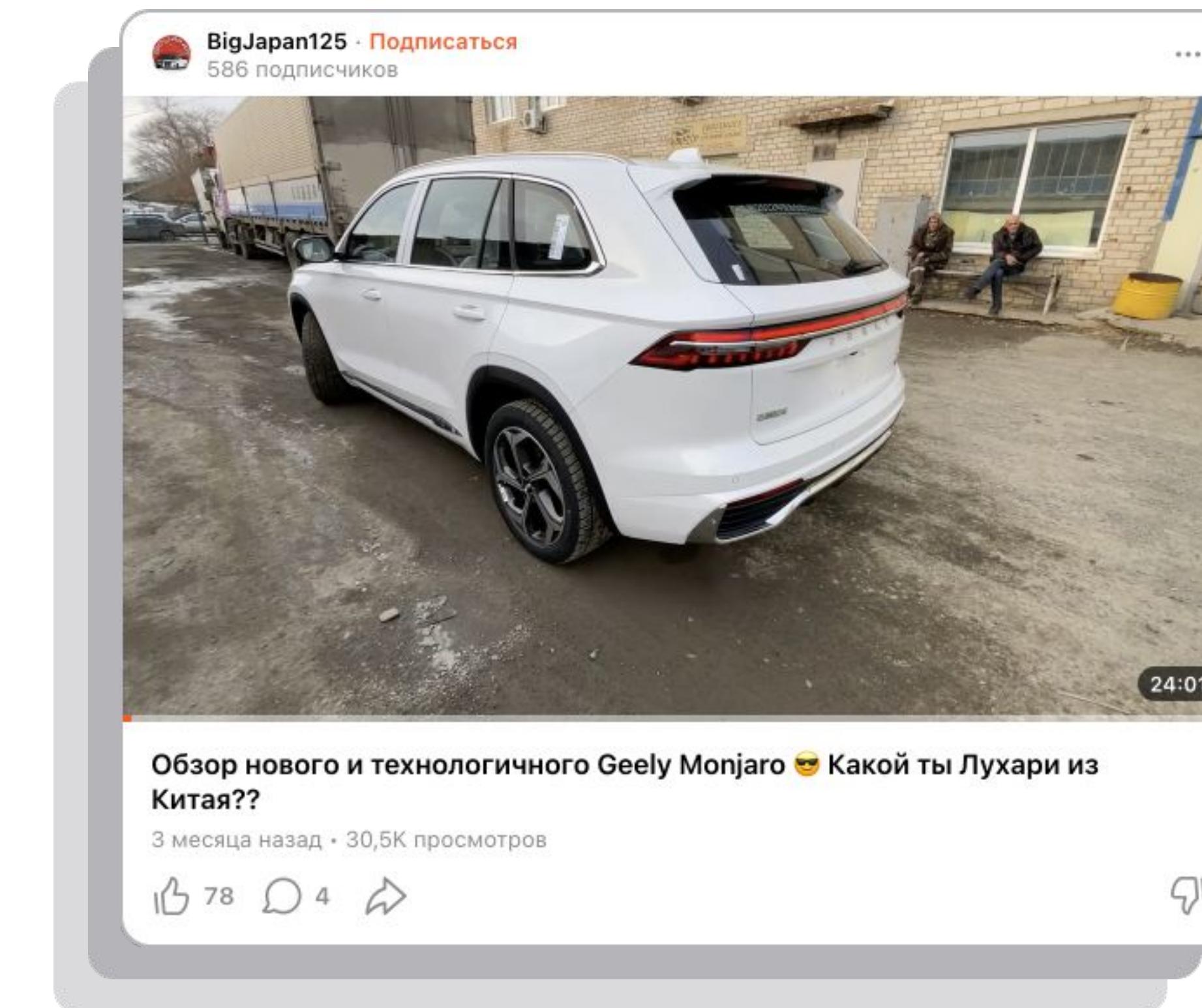
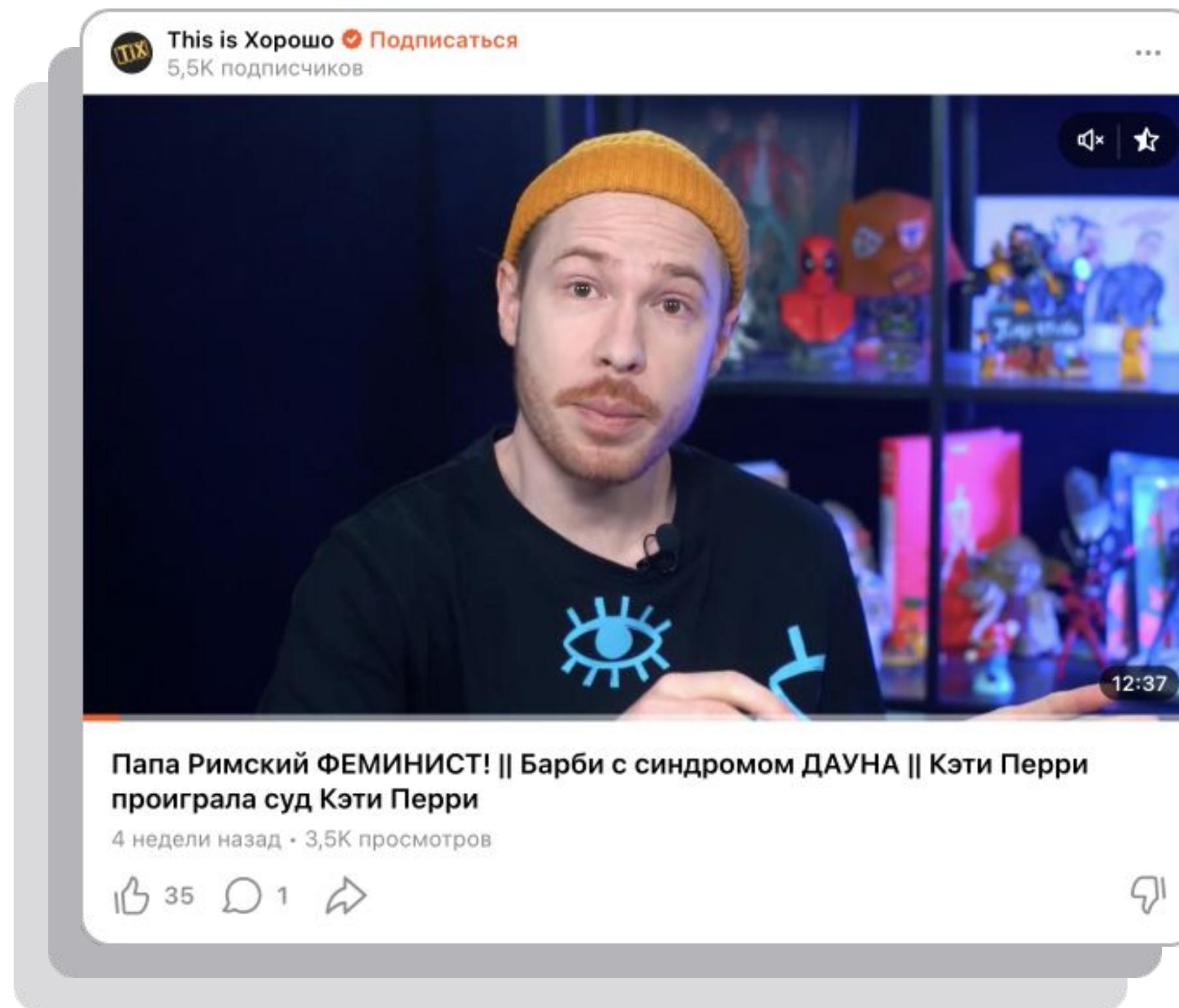


Saint
HighLoad++
2023



Как понять, что два айтема похожи?

Разметка похожестей краудом



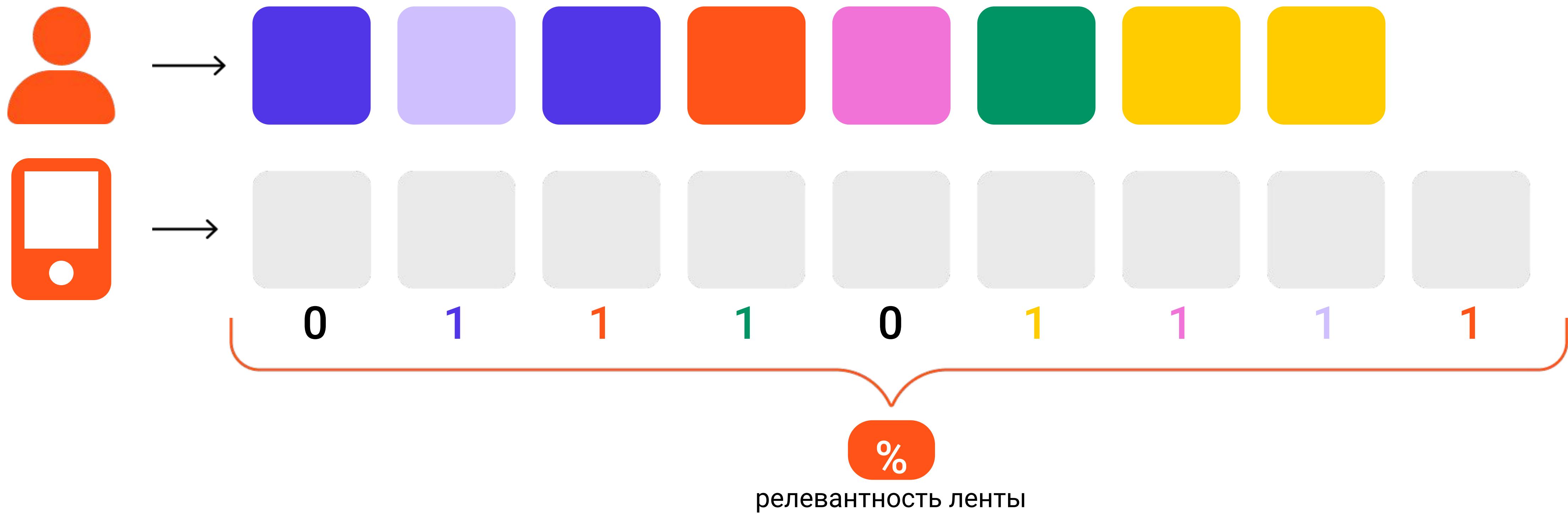
Дзен

Saint
HighLoad++
2023



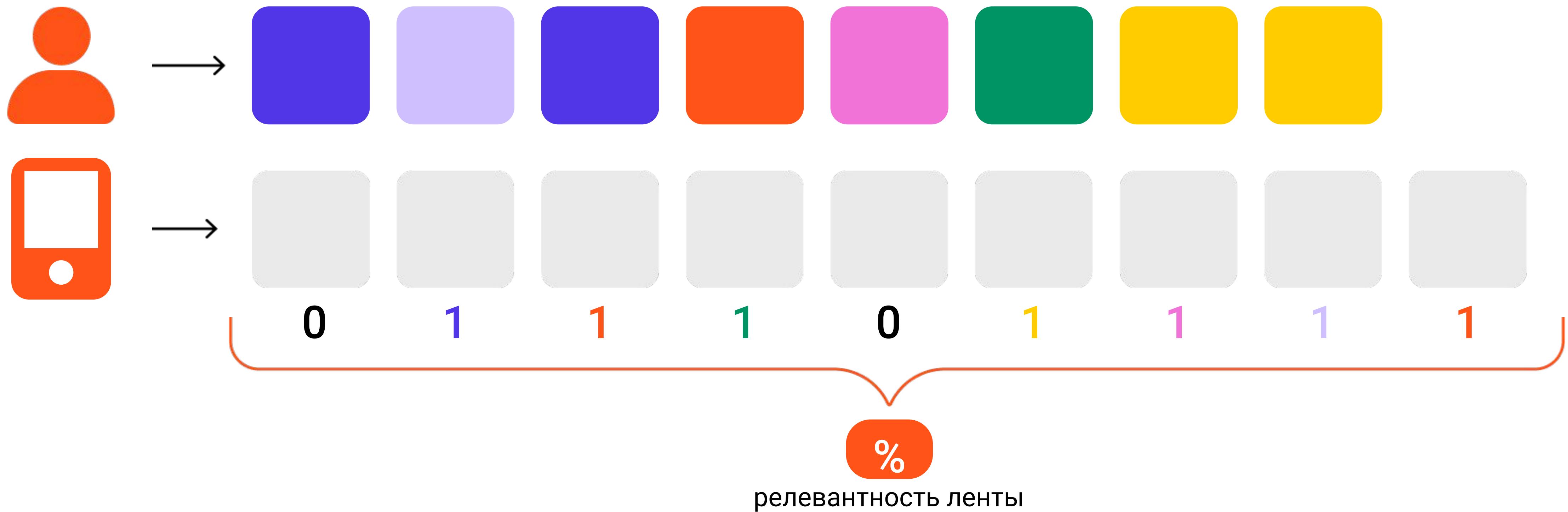
Проблемы

- ★ Часть ленты совсем не похожа на историю пользователя



Проблемы

- ❖ Часть ленты совсем не похожа на историю пользователя
- ❖ Результат замера Дзена для горячих пользователей:



Что хотим?

- ① Сделать этап генерации кандидатов простым
- ② 90% показов должны быть похожи на историю пользователя

Проблемы

Не все интересы представлены в ленте

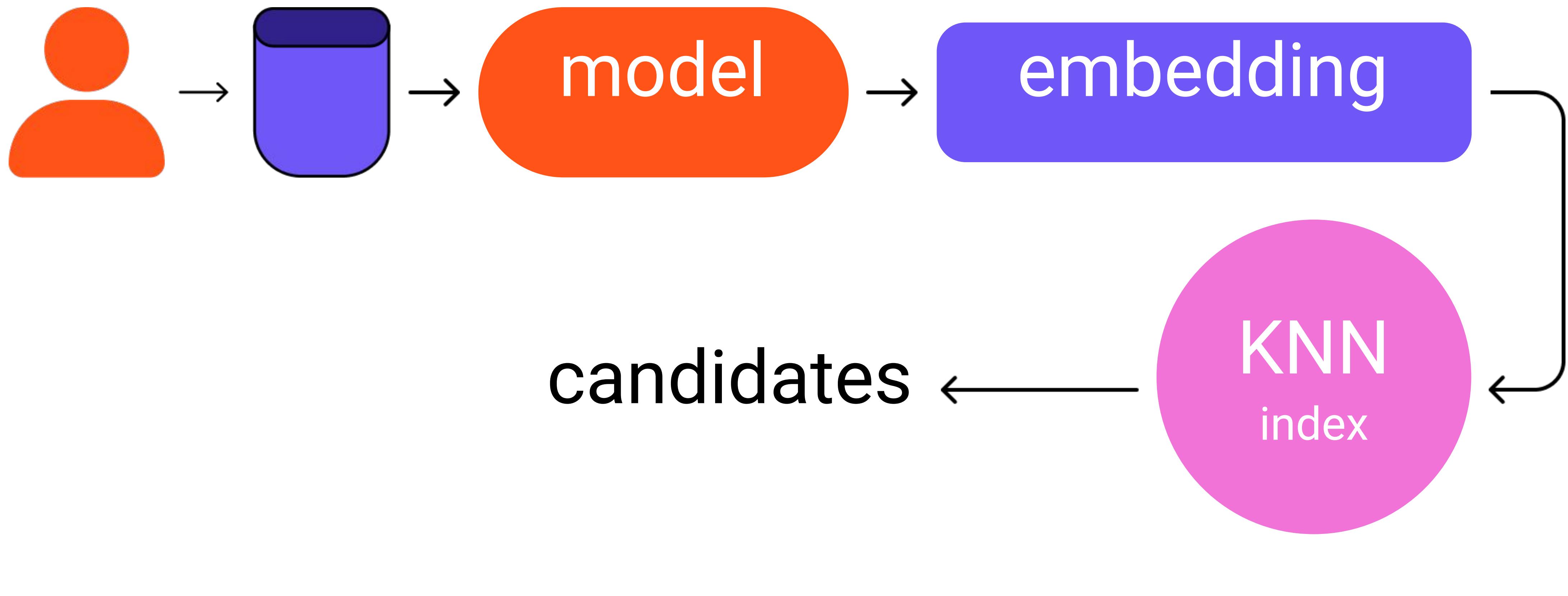


Saint
HighLoad++
2023



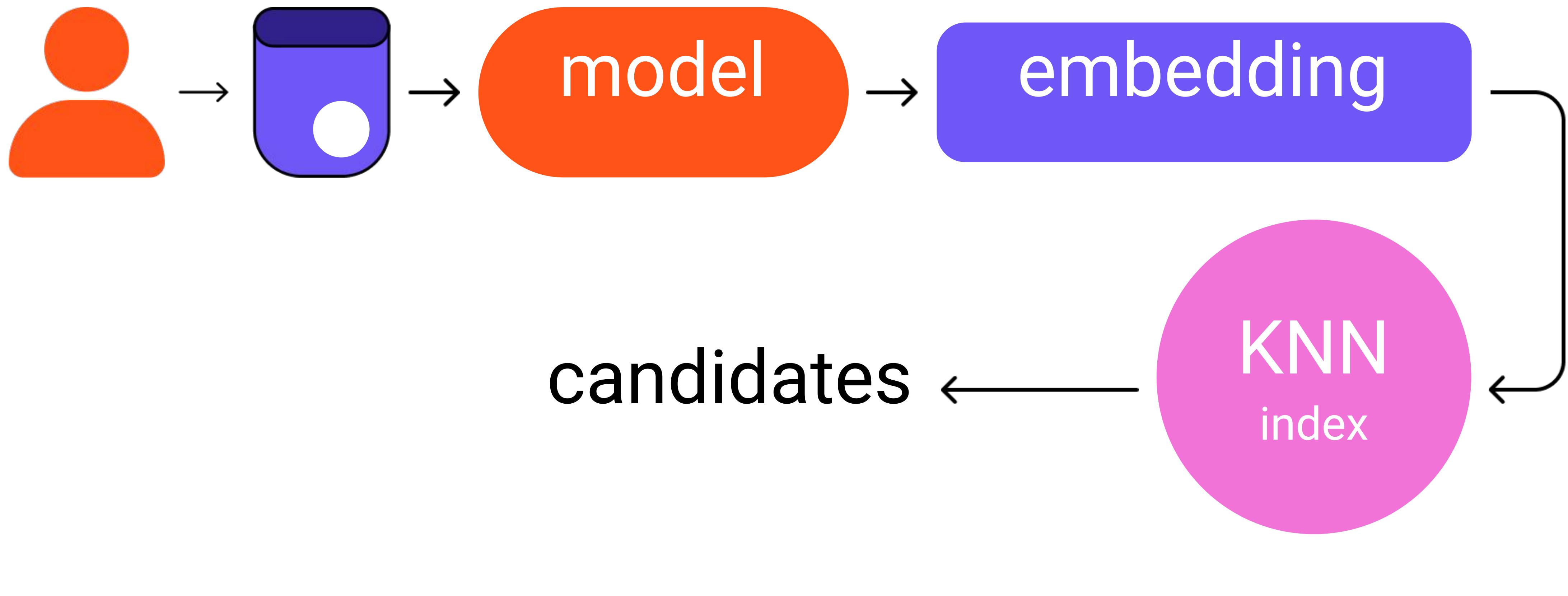
Проблемы

Не все интересы представлены в ленте



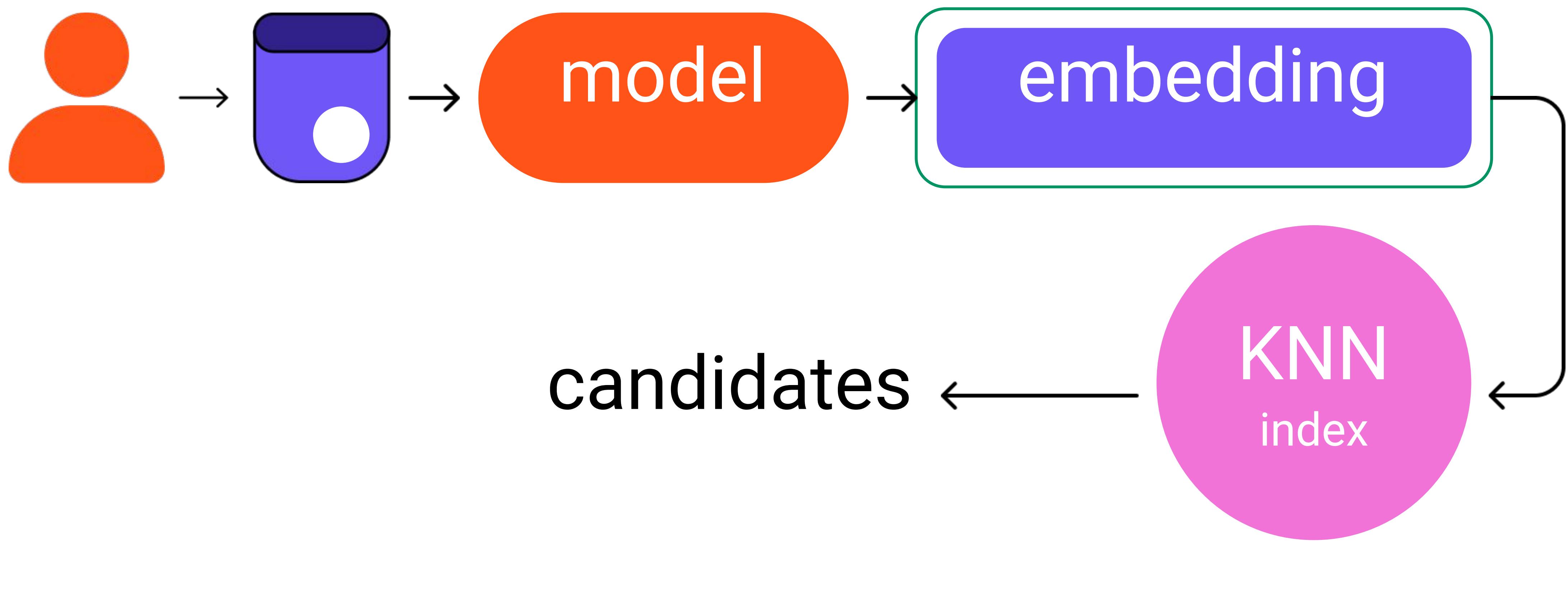
Проблемы

Не все интересы представлены в ленте



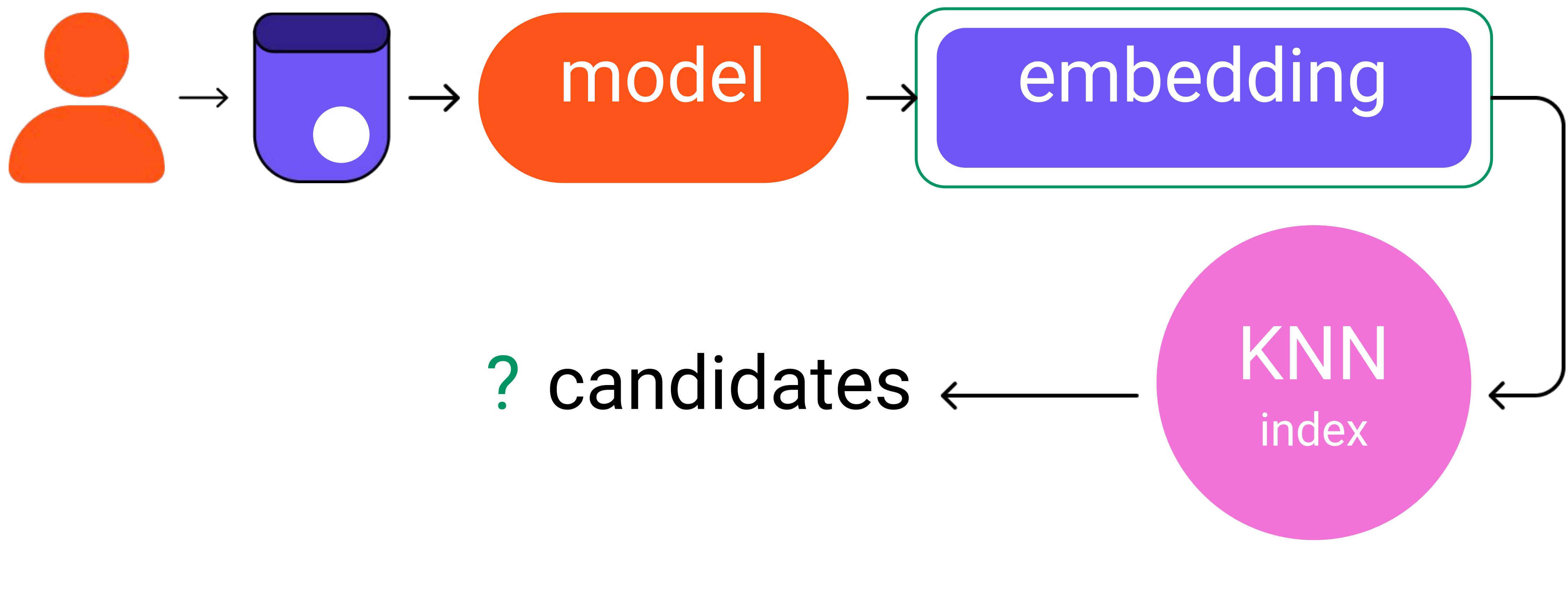
Проблемы

Не все интересы представлены в ленте



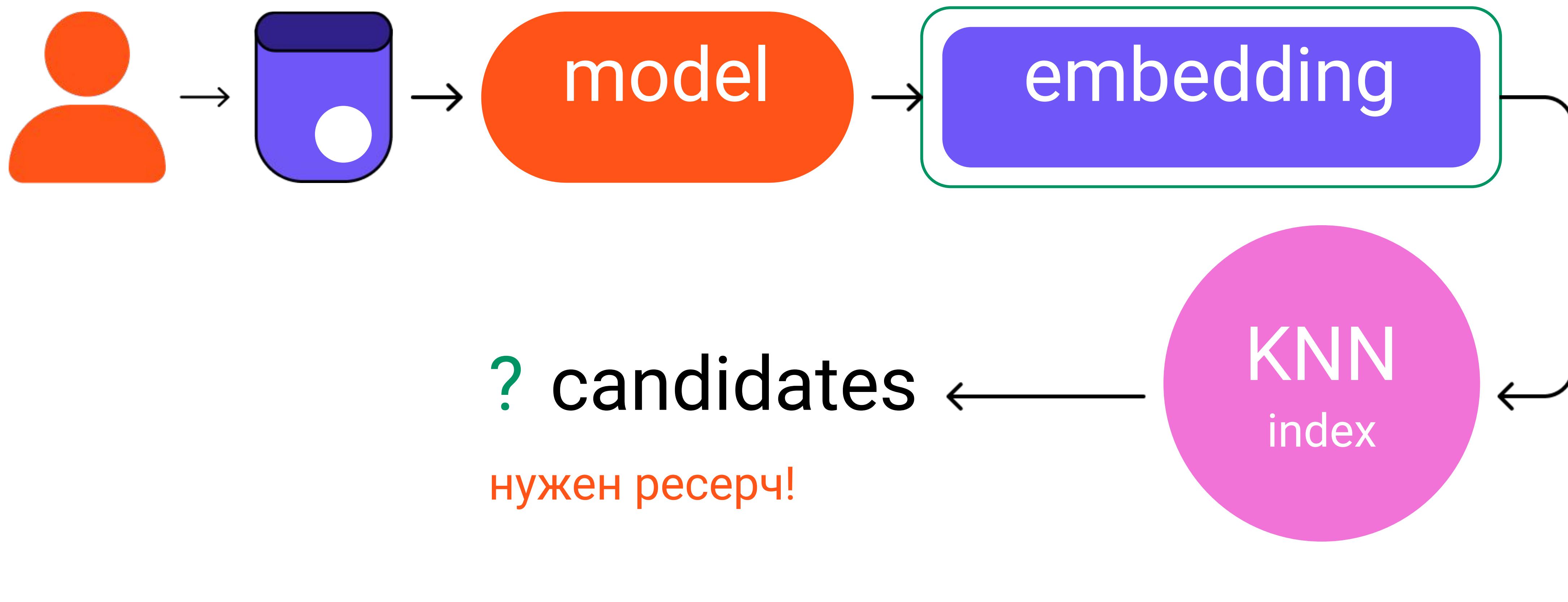
Проблемы

Не все интересы представлены в ленте



Проблемы

Не все интересы представлены в ленте



Что хотим?

- ① Сделать этап генерации кандидатов простым
- ② 90% показов должны быть похожи на историю пользователя
- ③ Гарантировать использование всей истории пользователя



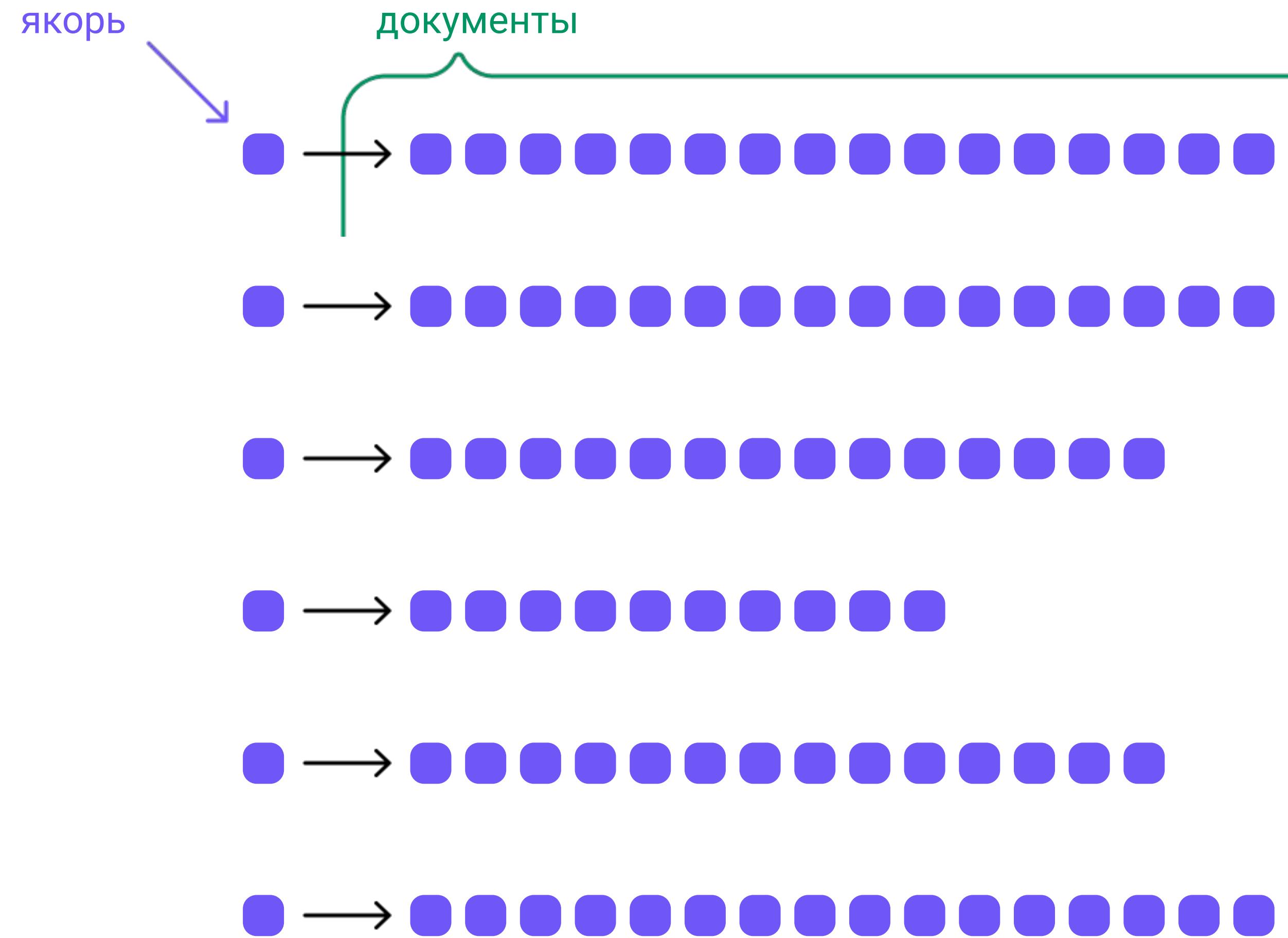
Saint
HighLoad++
2023



item2item схема рекомендаций



Item2Item

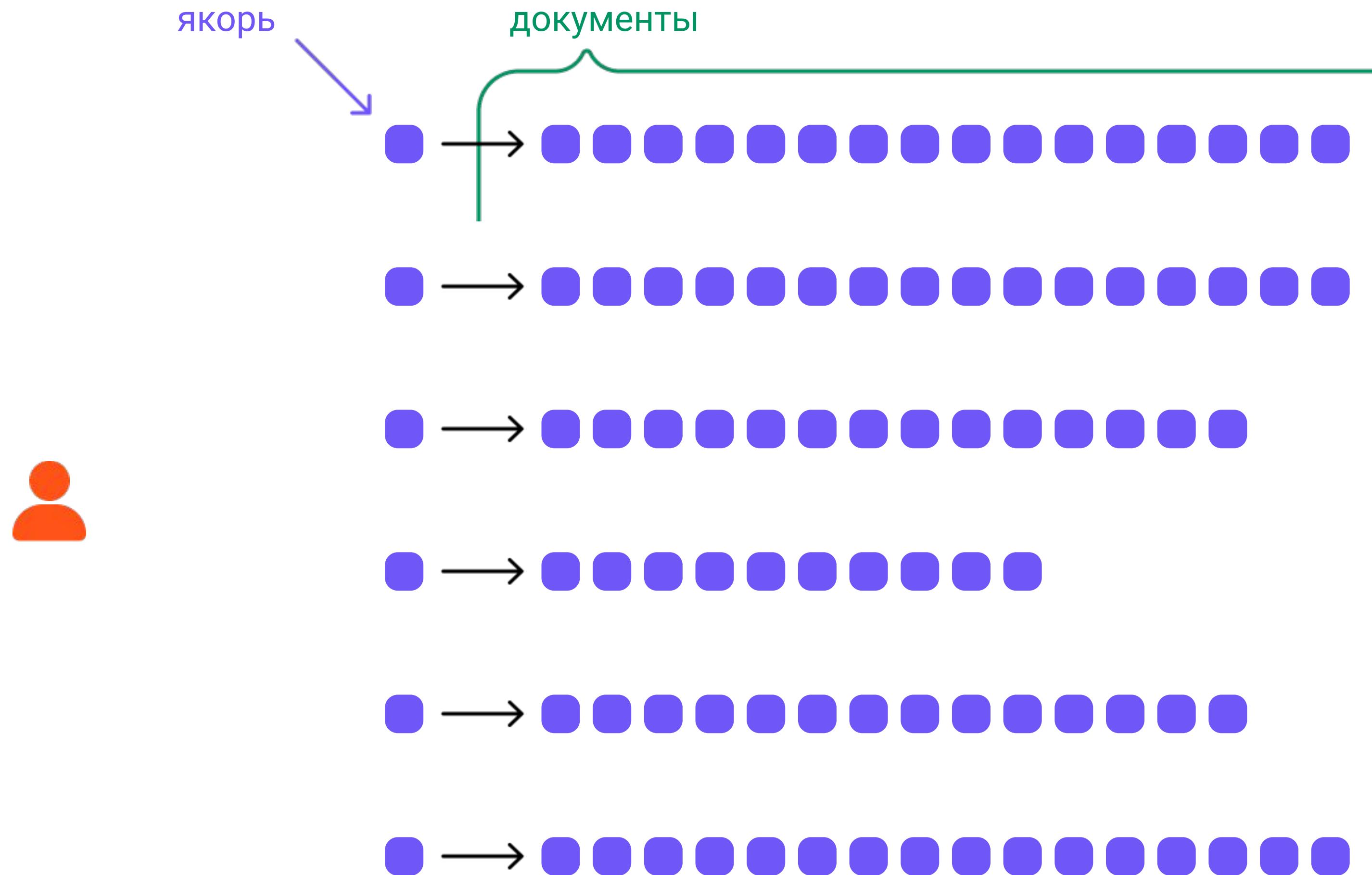


Дзен

Saint
HighLoad++
2023



Item2Item

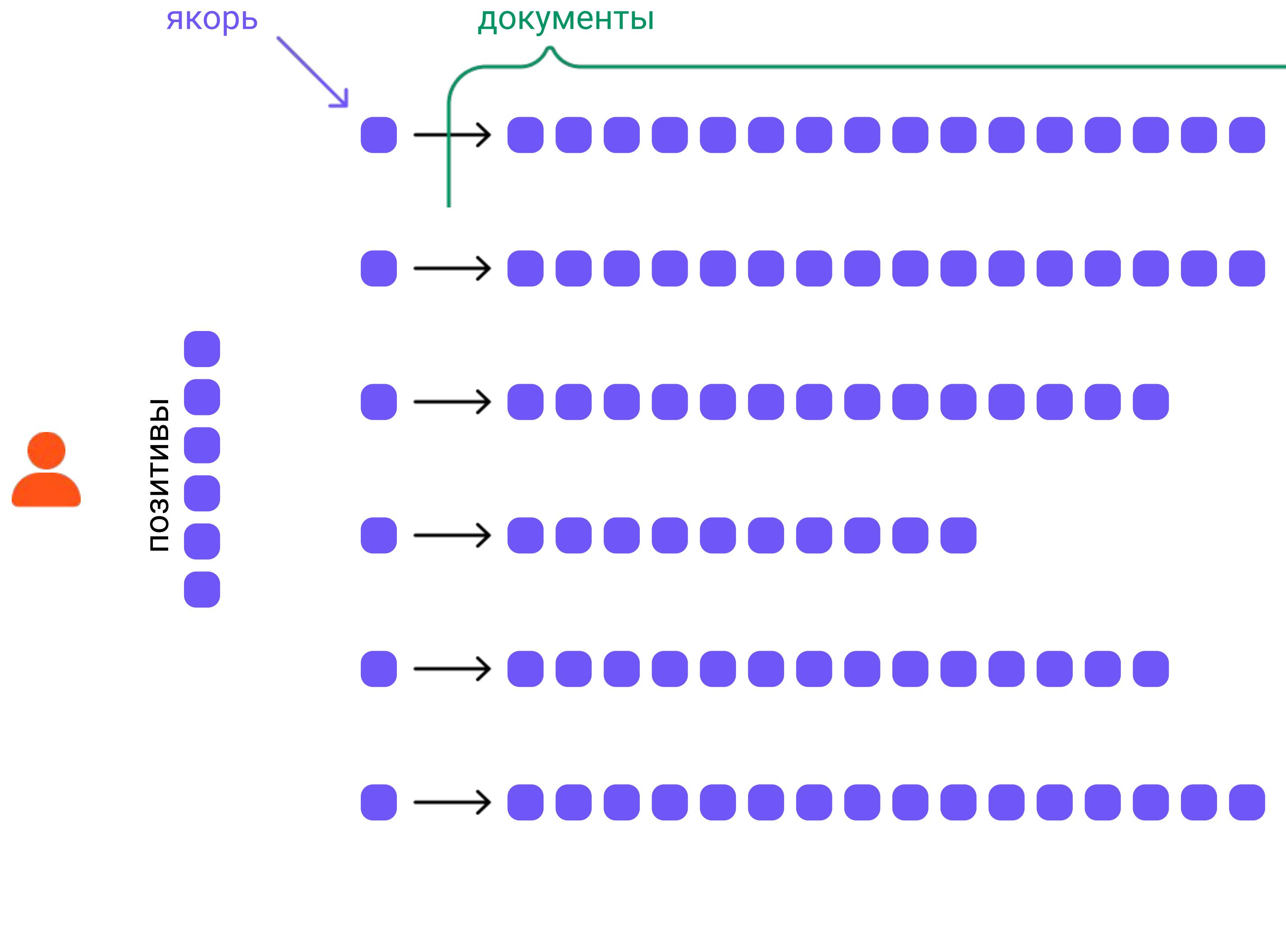


Дзен

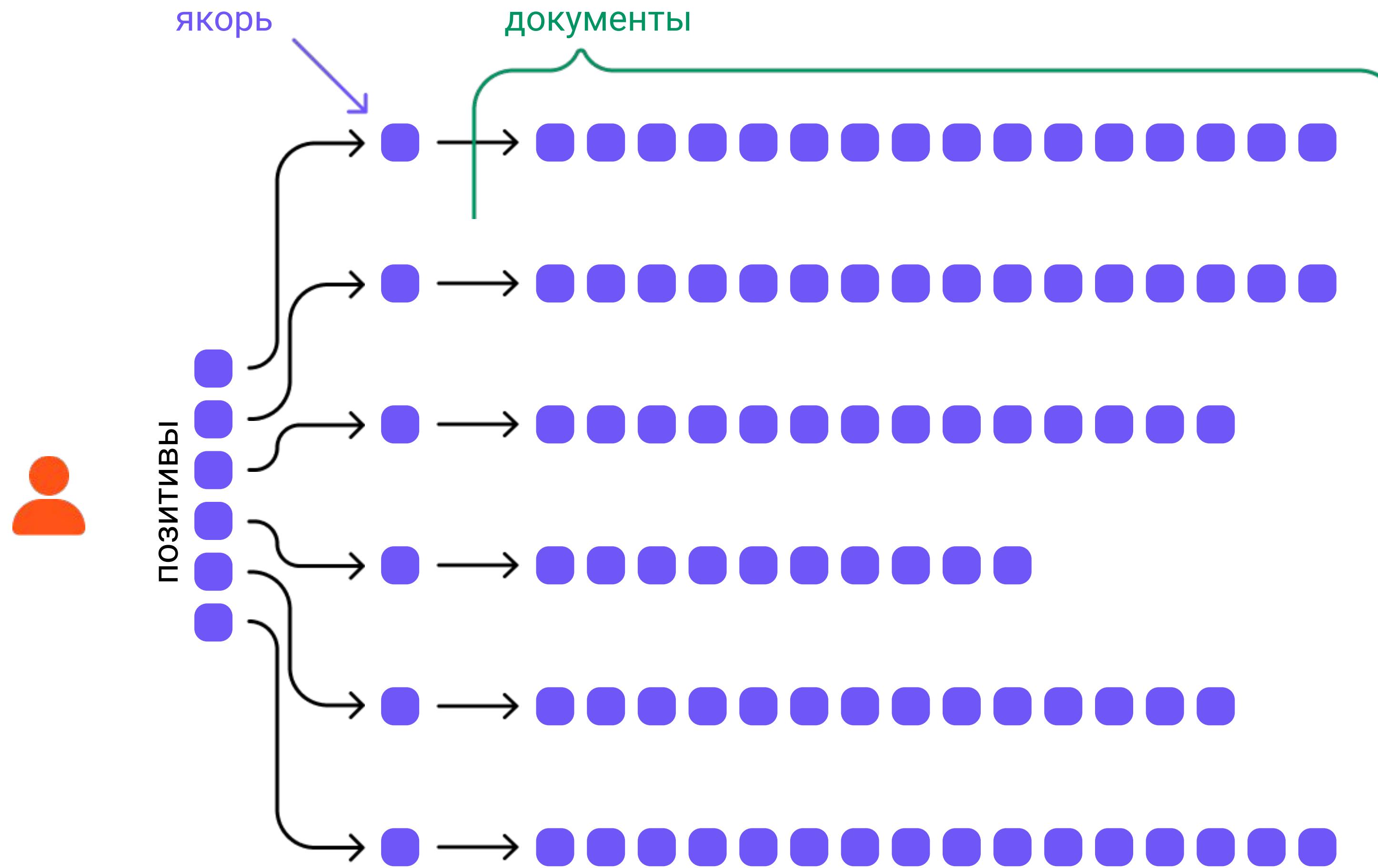
Saint
HighLoad++
2023



Item2Item



Item2Item

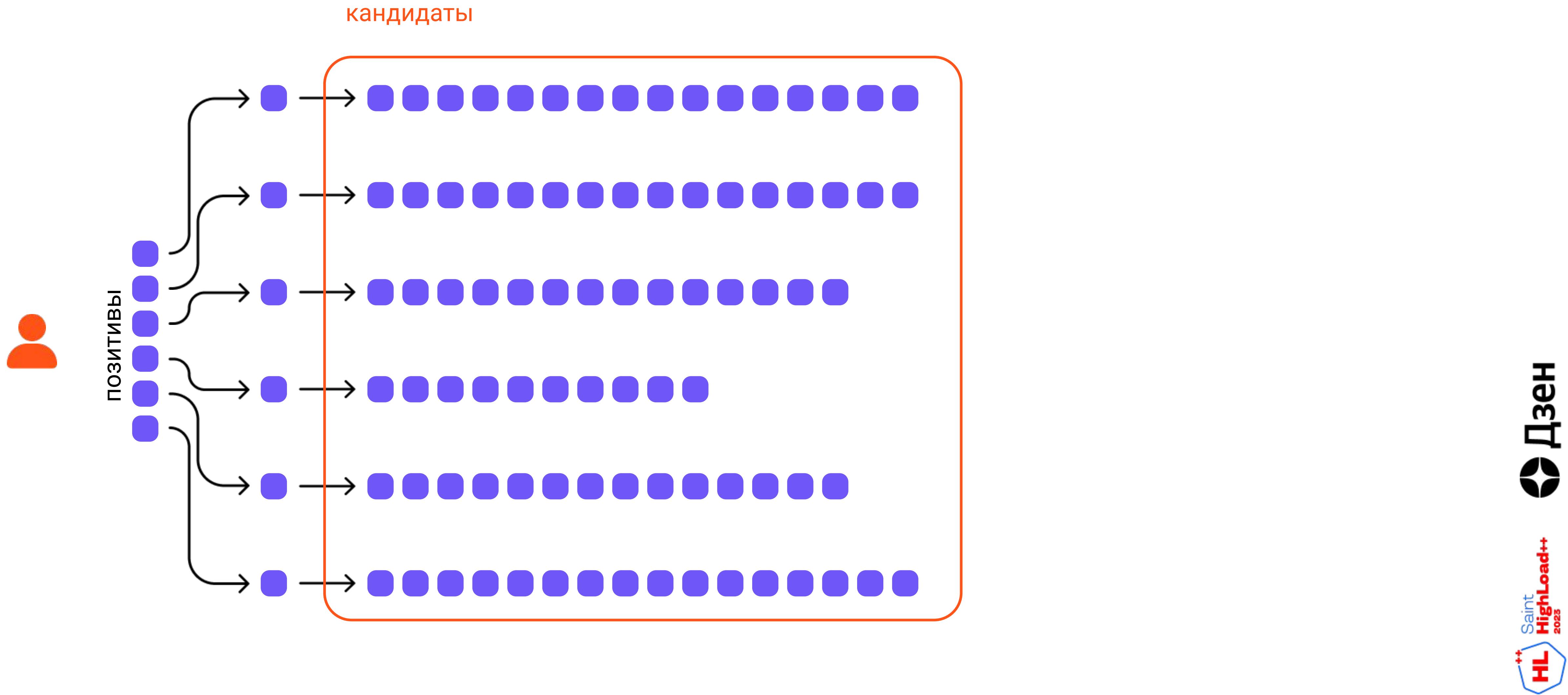


Дзен

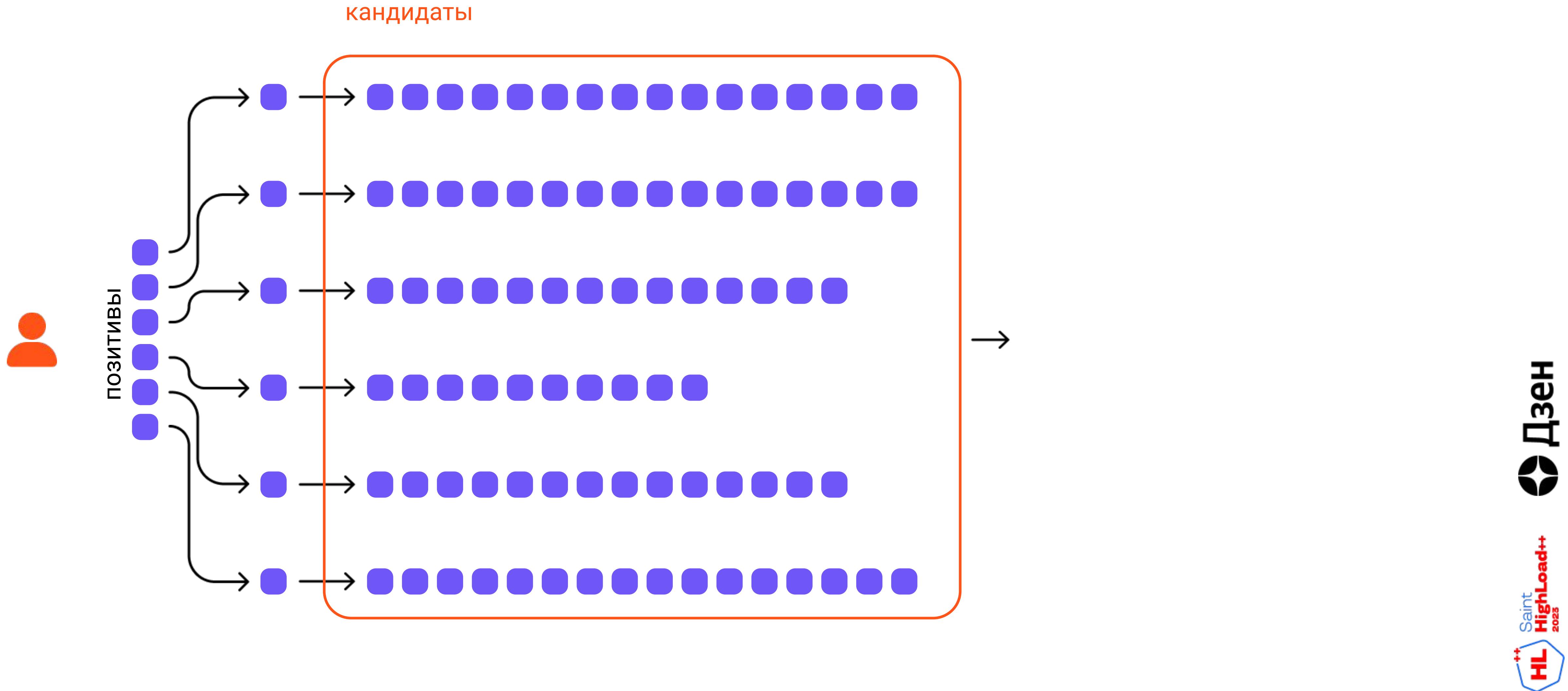
Saint
HighLoad++
2023



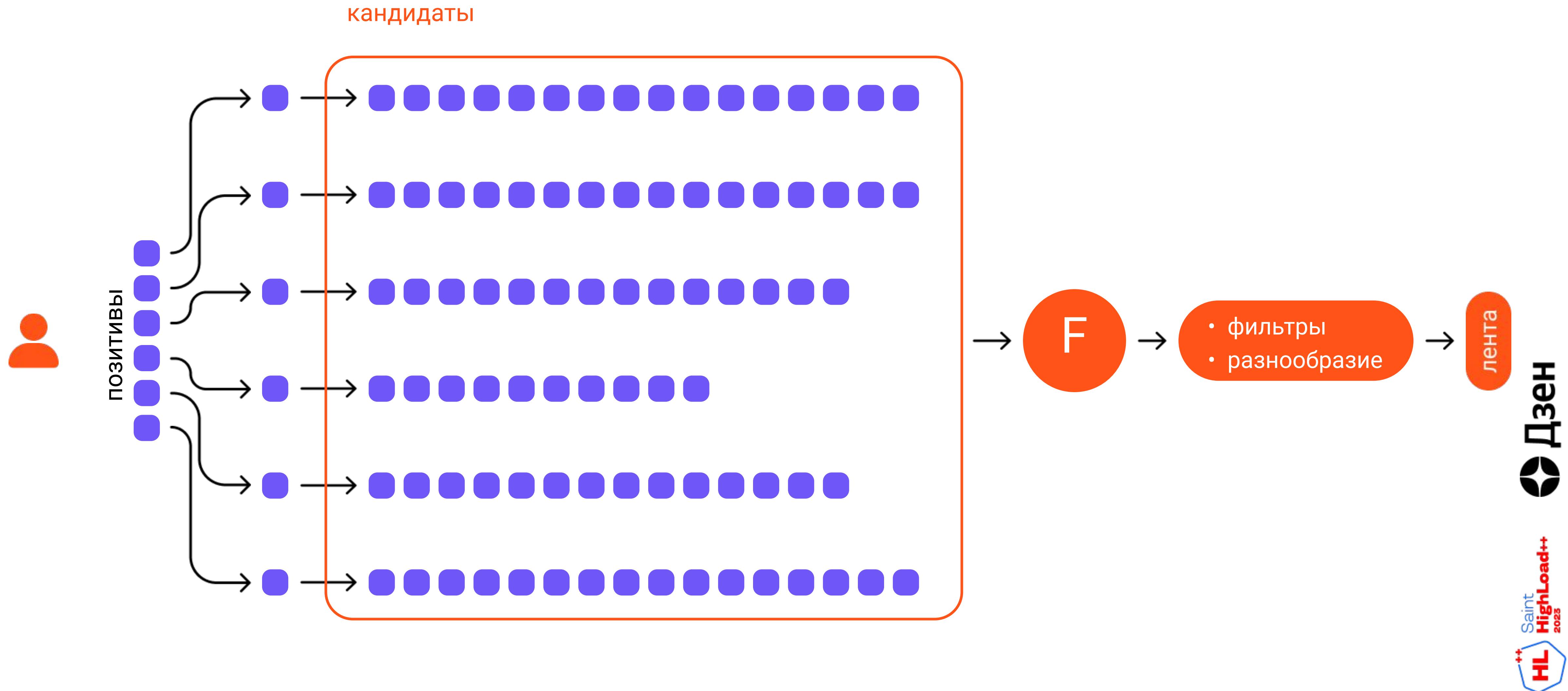
Item2Item



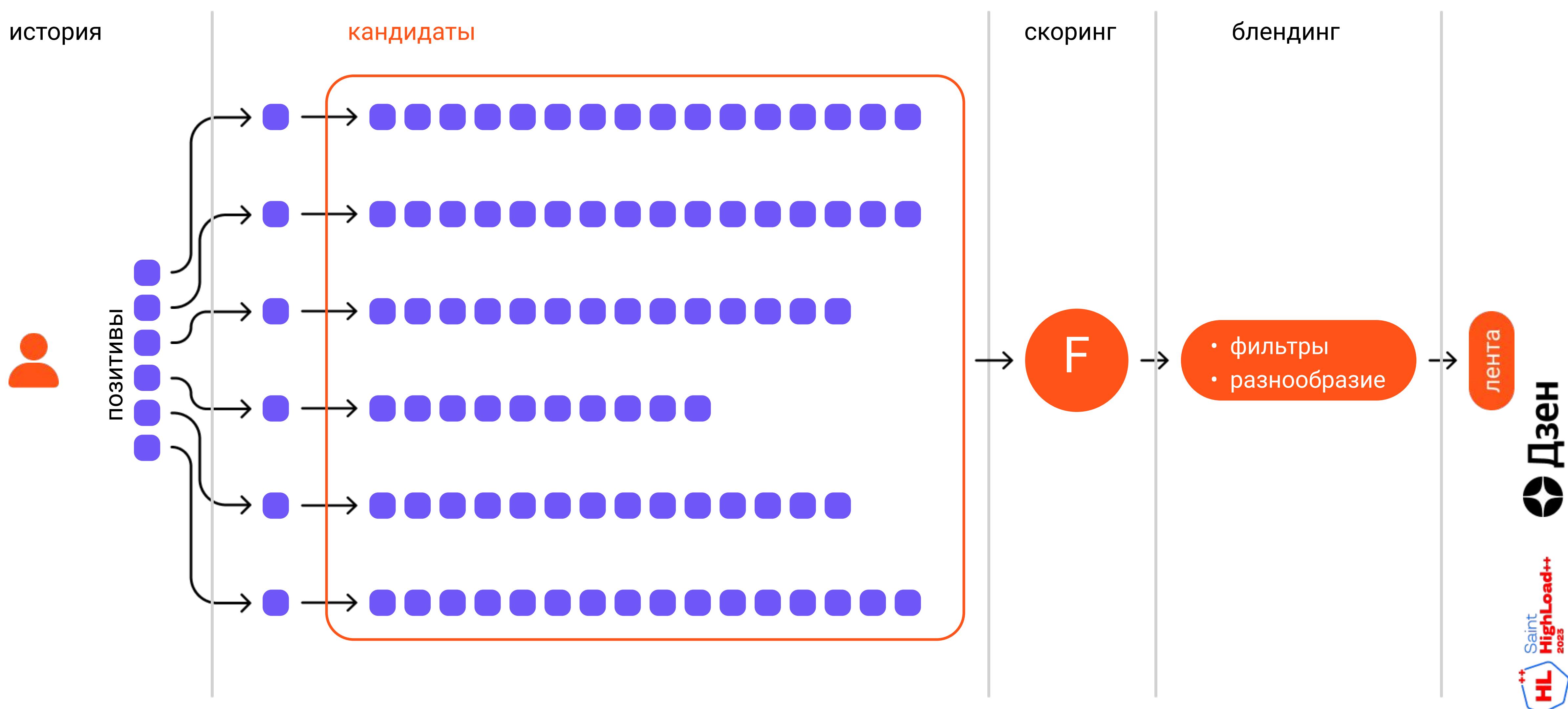
Item2Item



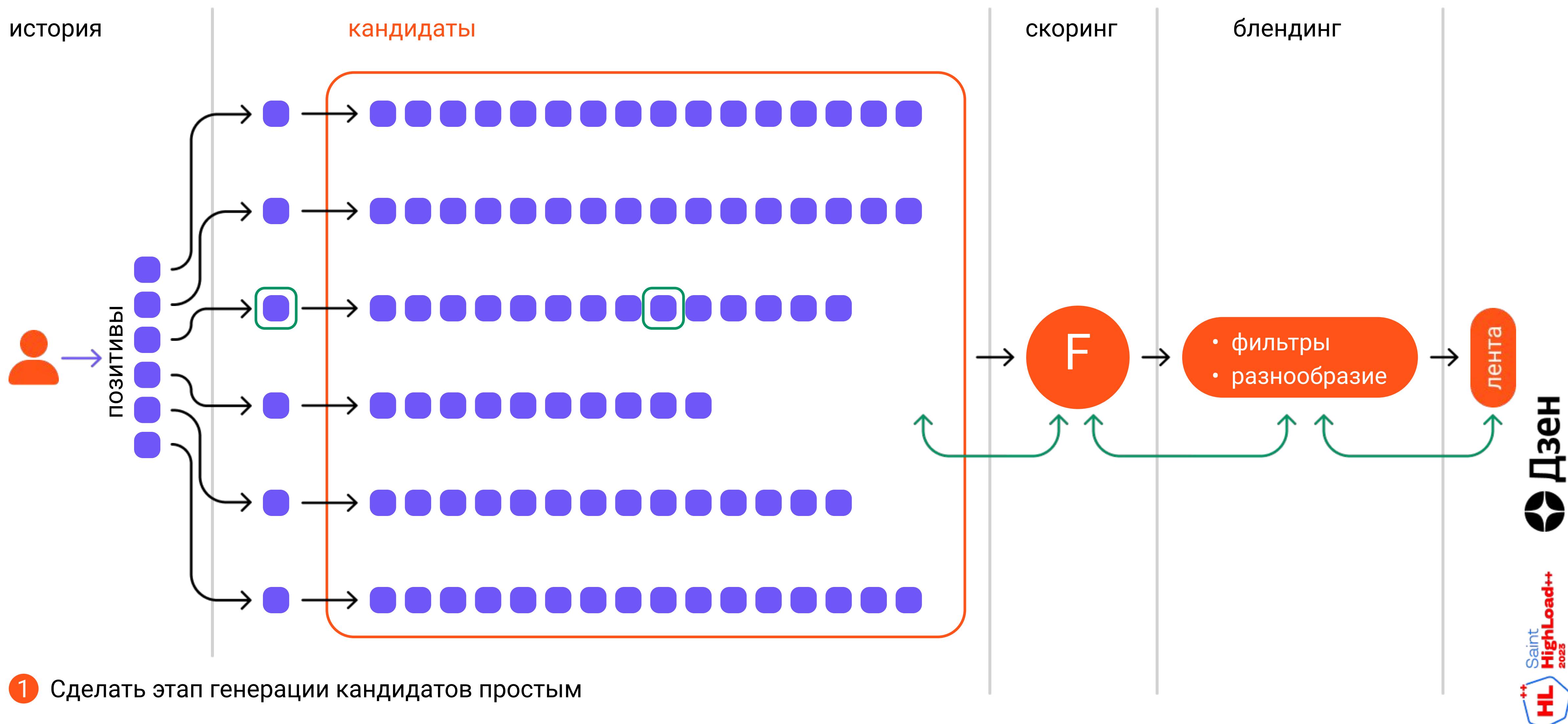
Item2Item



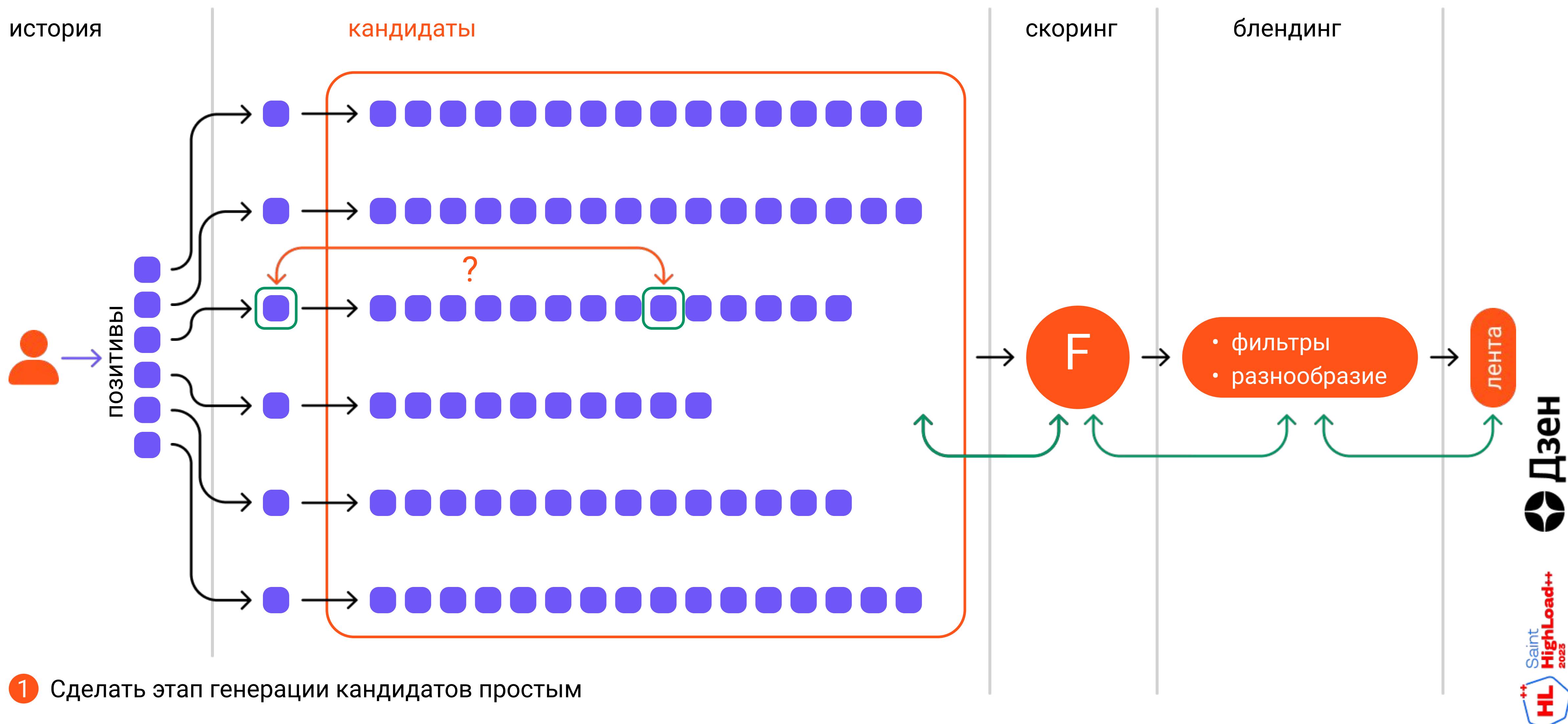
Item2Item



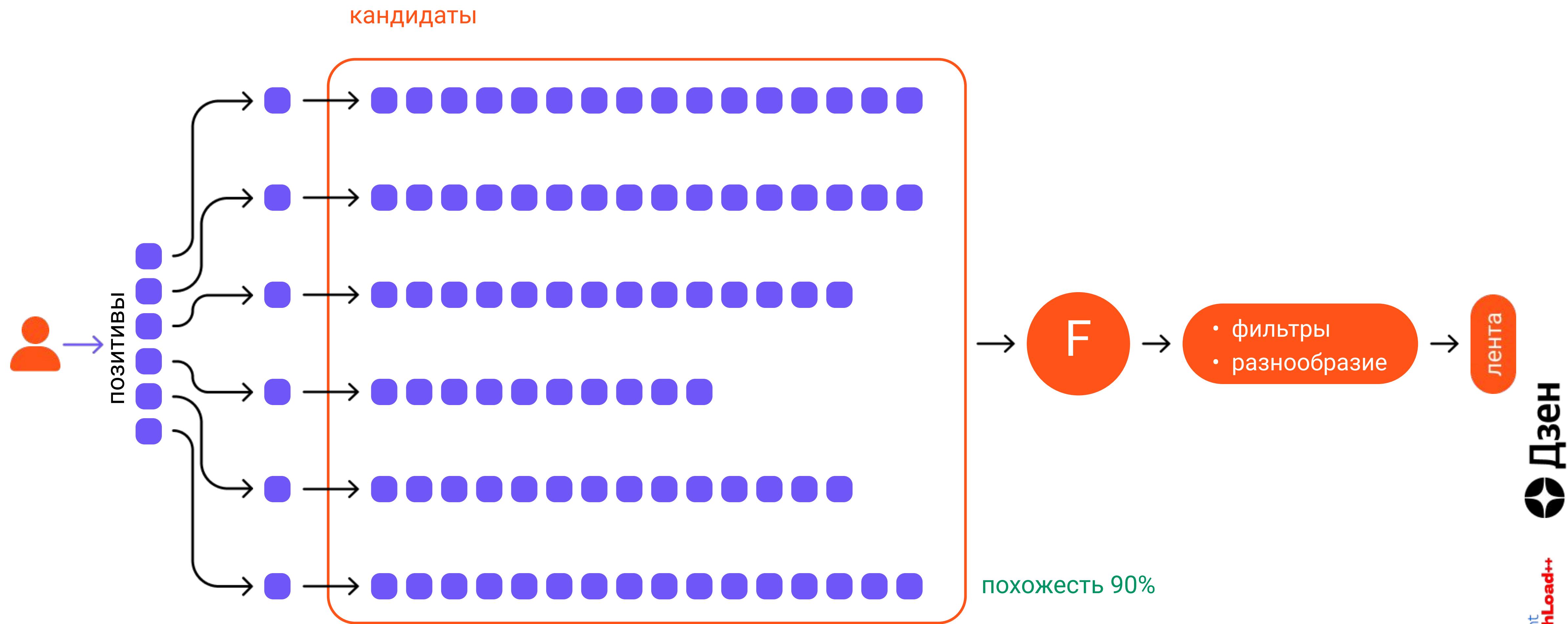
Item2Item



Item2Item

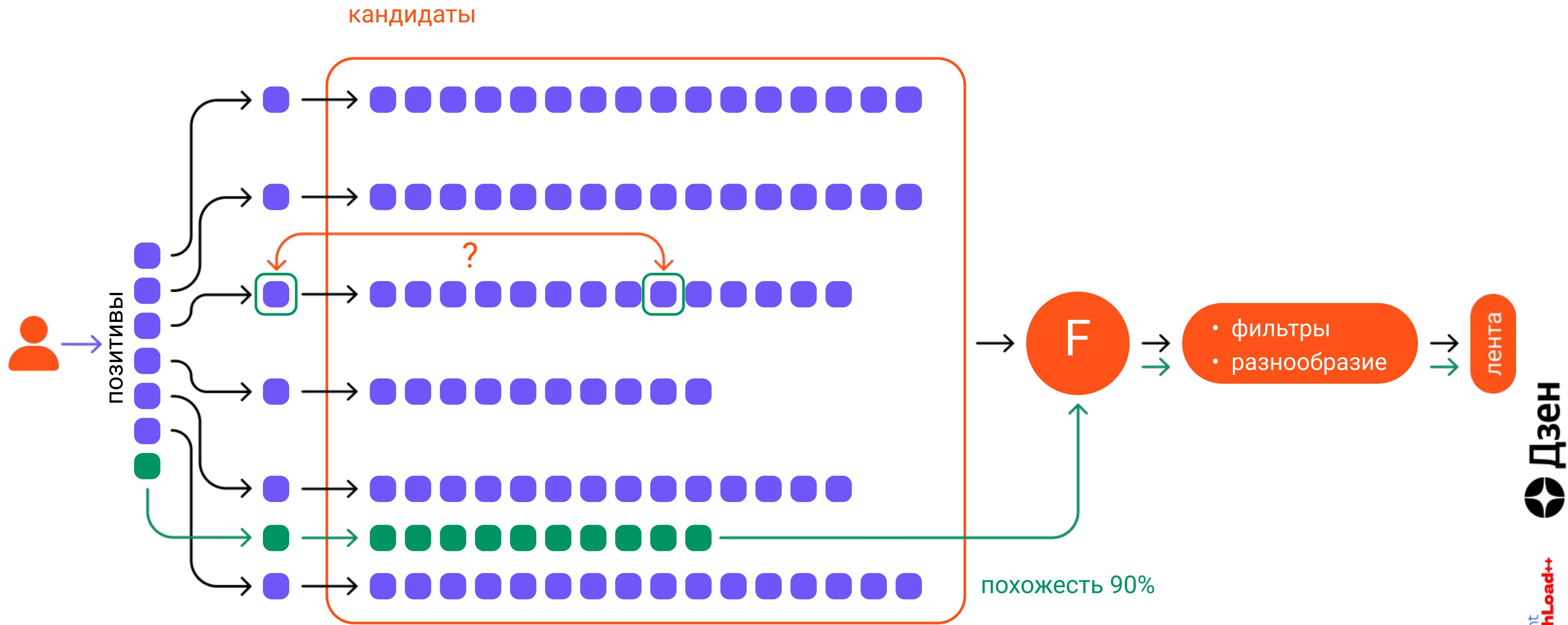


Item2Item



② Сделать ленту похожей на историю юзера

Item2Item



③ Гарантировать использование всей истории пользователя

Item2Item: как построить список

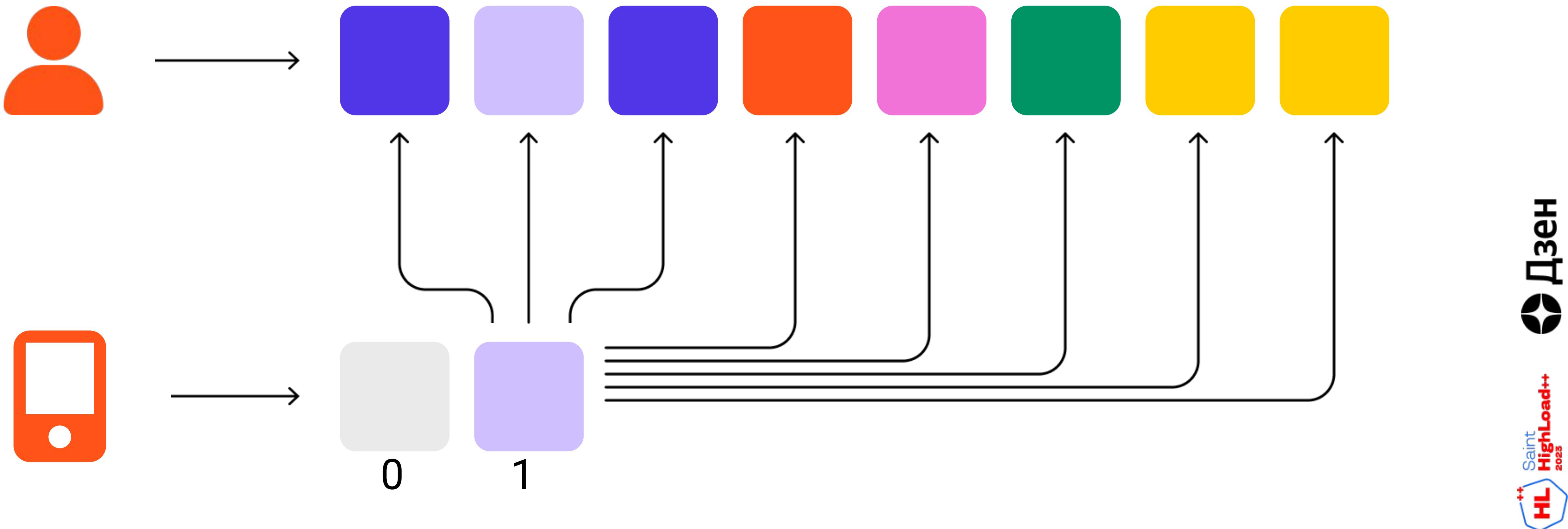
К айтому A составить список из 90% релевантных айтемов

Решение

- ❖ классификатор релевантности $f(A, item)$
- ❖ таргет: ?

Проблемы

Часть ленты совсем не похожа на историю пользователя



Item2Item: как построить список

К айтему A составить список из 90% релевантных айтемов

Решение

- ❖ классификатор релевантности $f(A, item)$
- ❖ таргет: похожесть по разметке краудом

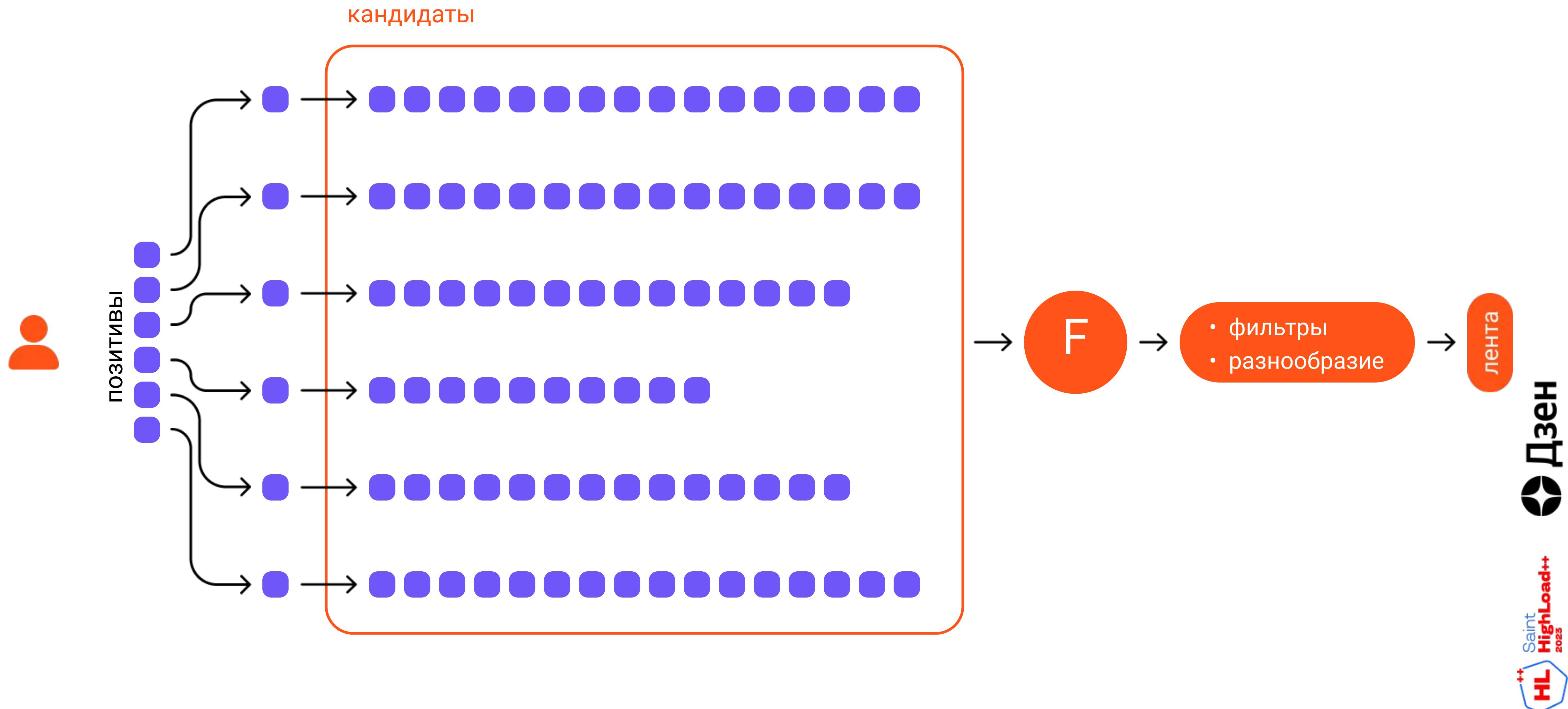
Item2Item: как построить список

К айтему A составить список из 90% релевантных айтемов

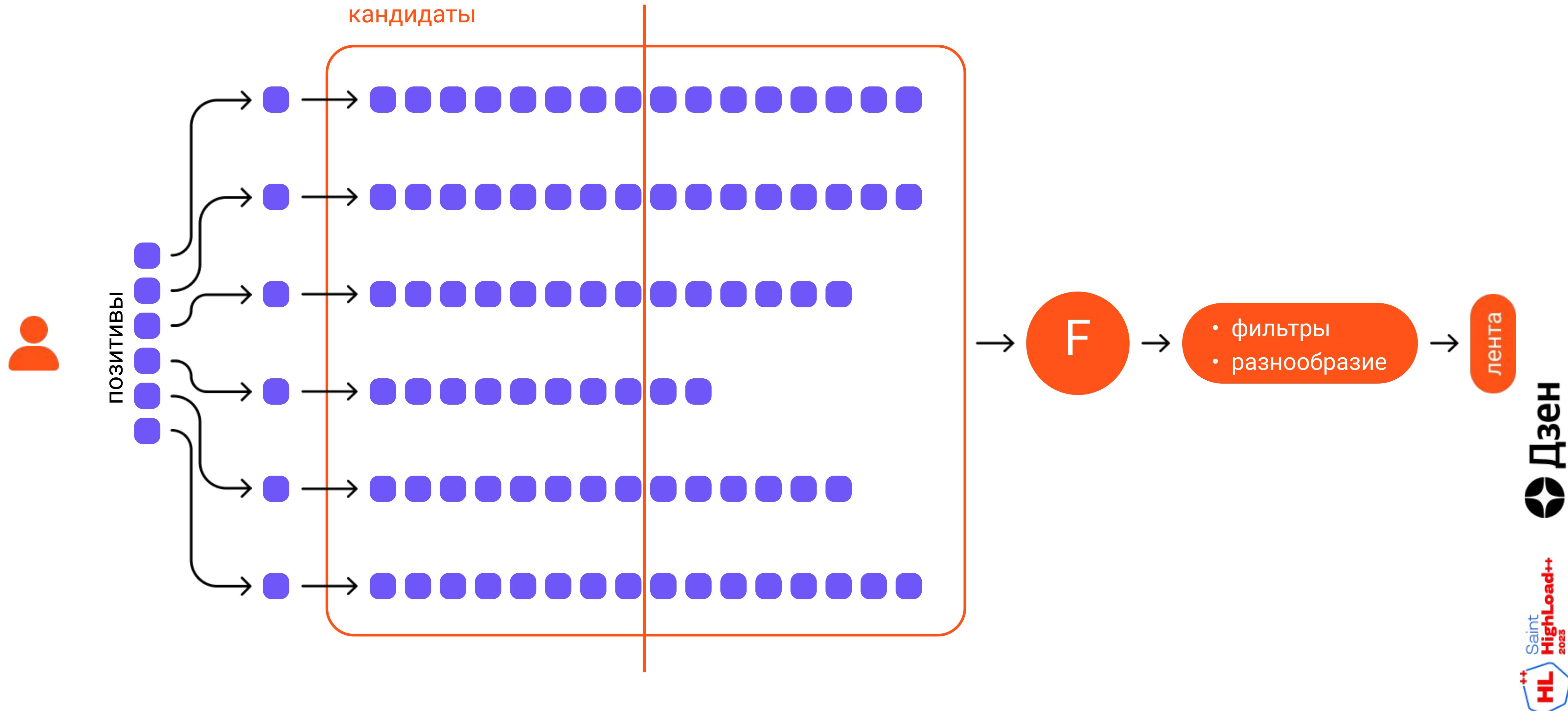
Решение

- ★ классификатор релевантности $f(A, item)$
- ★ таргет: похожесть по разметке краудом
- ★ признаки:
 - ★ контентная похожесть
 - ★ похожесть авторов
 - ★ информация о документах
- ★ при построении списка поскорить всю контентную базу

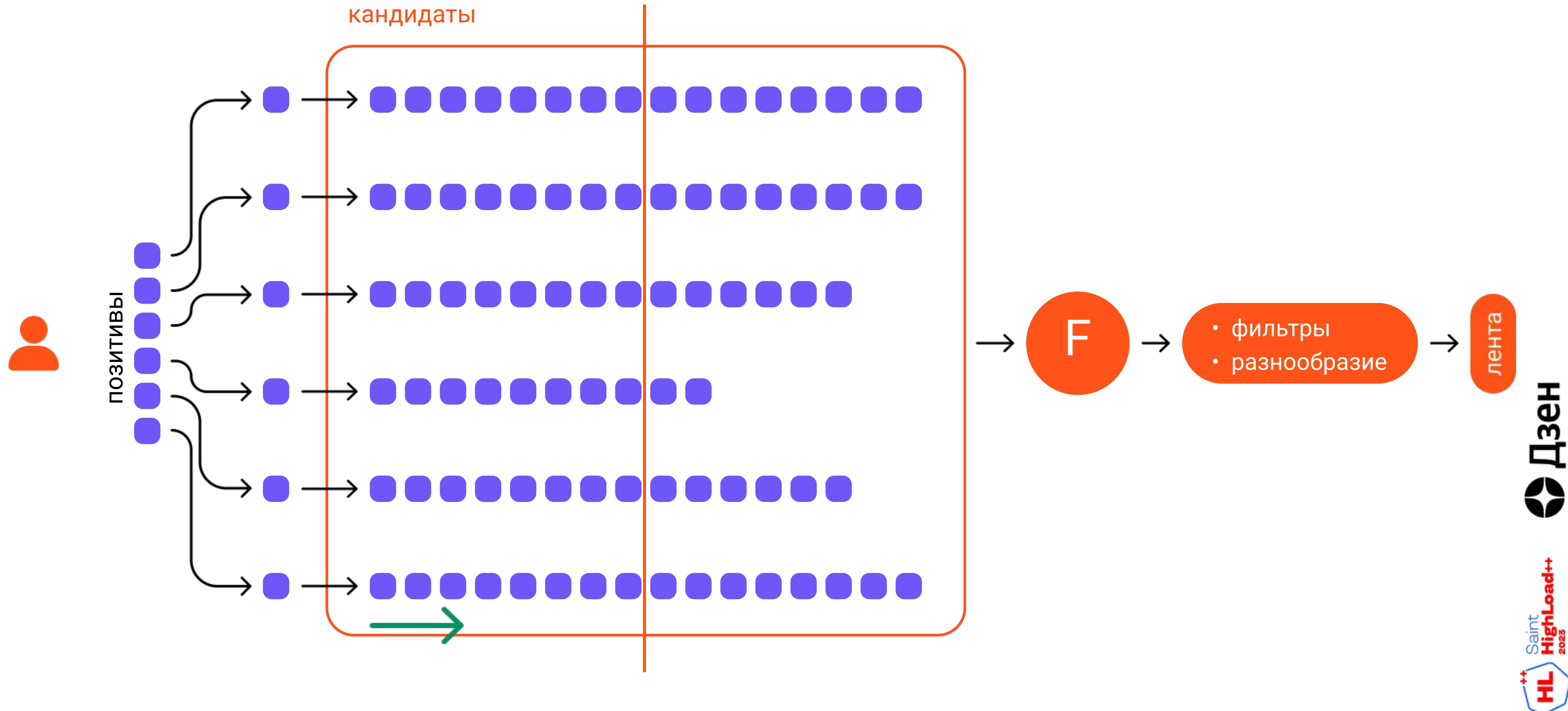
Item2Item



Item2Item



Item2Item



Item2Item: как построить список

Айтемы в списке нужно отранжировать

Решение 1

- отранжируем по «условному ctr»

A – позитив

C

–

+

B

–

+



Item2Item: как построить список

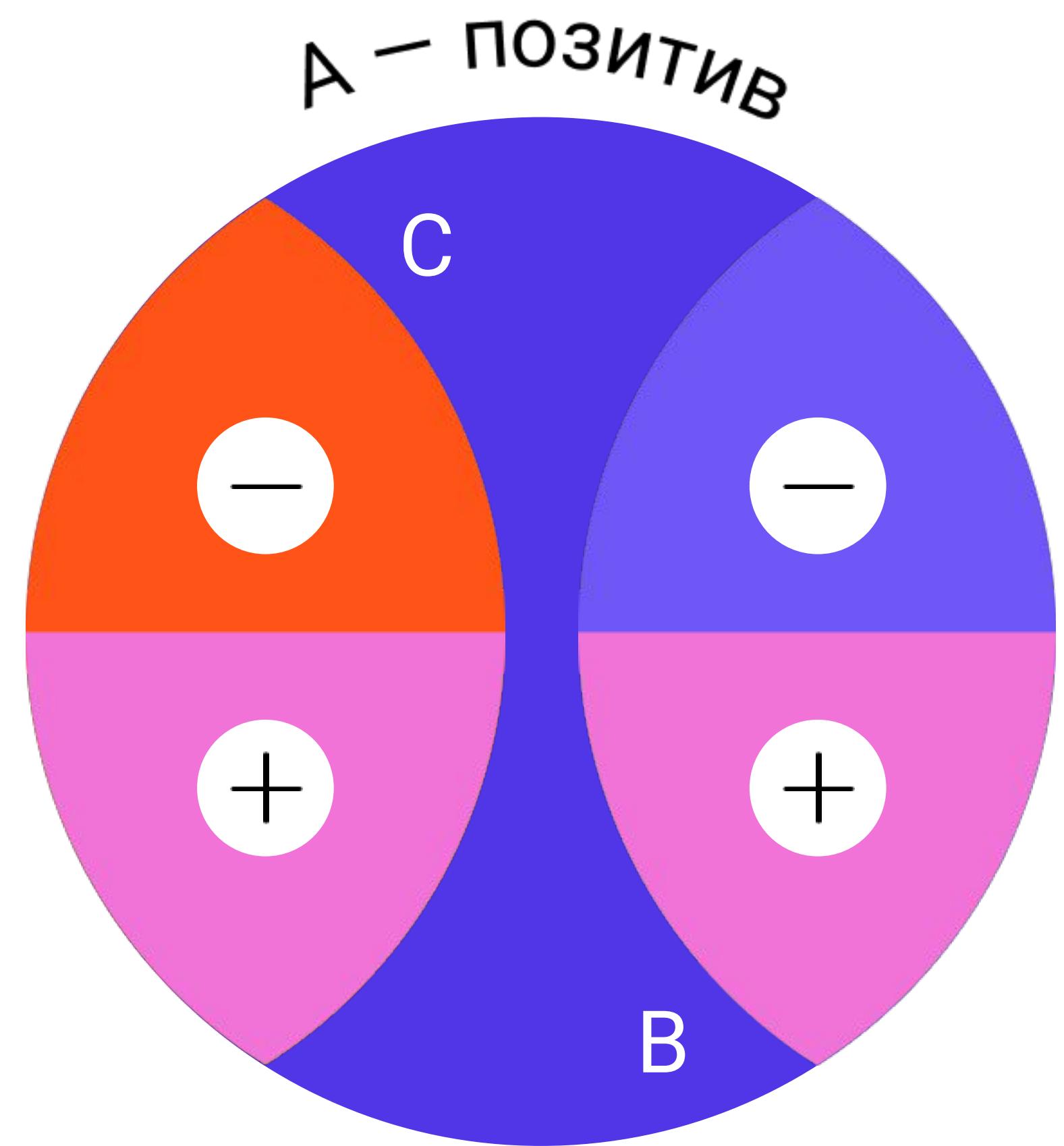
Айтемы в списке нужно отранжировать

Решение 1

- ★ отранжируем по «условному ctr»

Плюсы

- ★ это работает



Item2Item: как построить список

Айтемы в списке нужно отранжировать

Решение 1

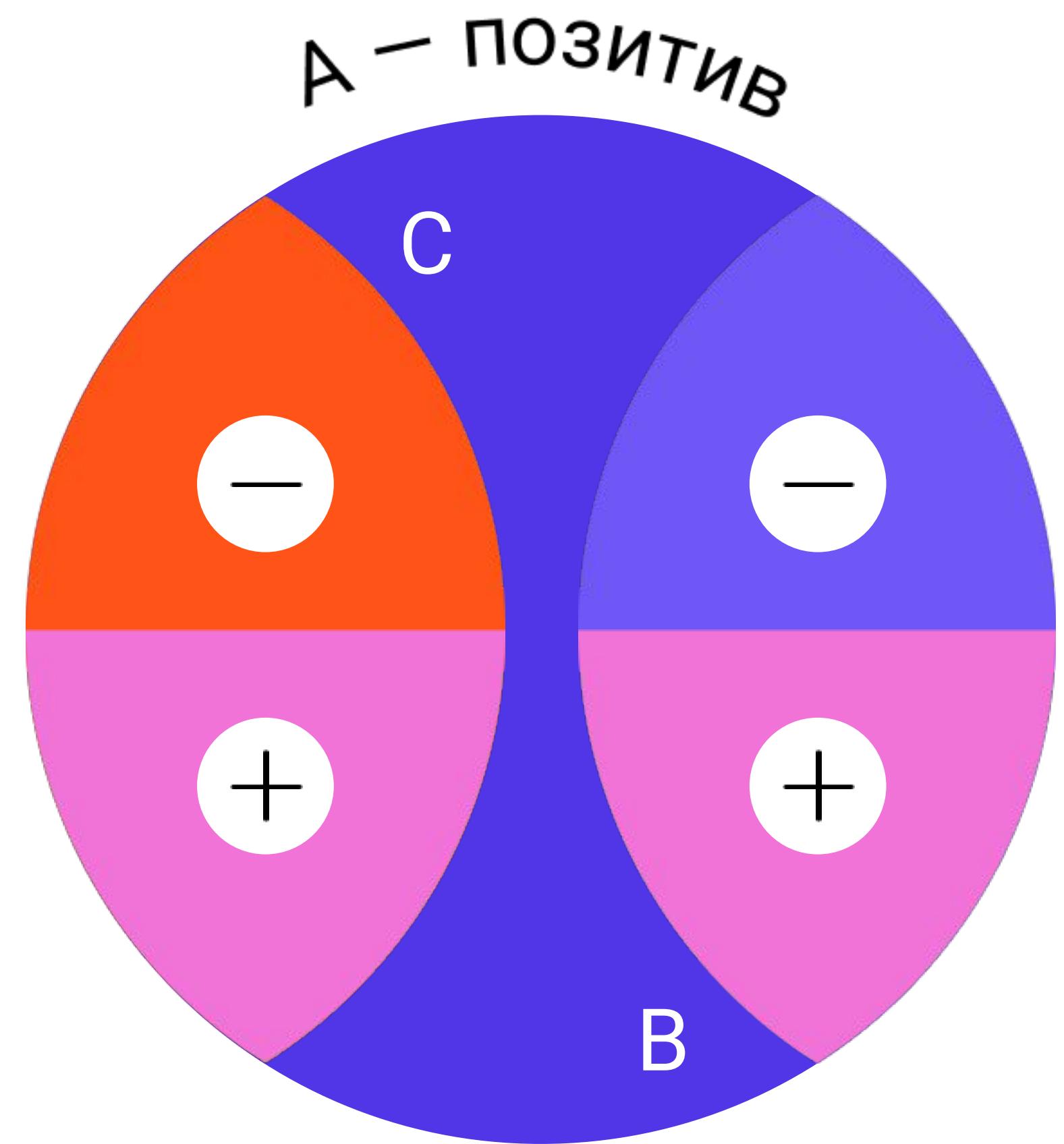
- ❖ отранжируем по «условному ctr»

Плюсы

- ❖ это работает

Минусы

- ❖ данные разреженные
- ❖ не работает на непрогретых айтемах

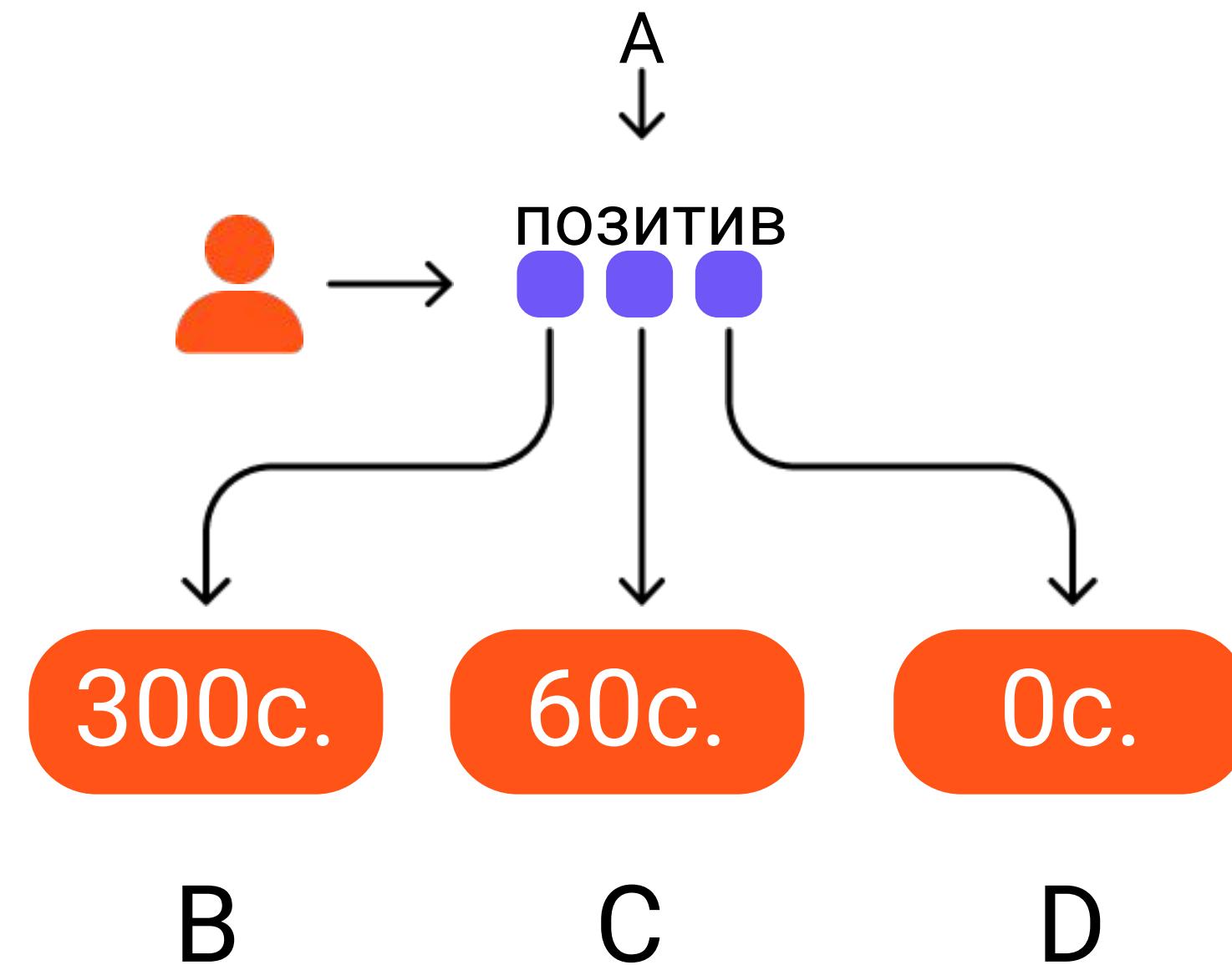


Item2Item: как построить список

Айтемы в списке нужно отранжировать

Решение 2

- ★ залогируем «позитив»

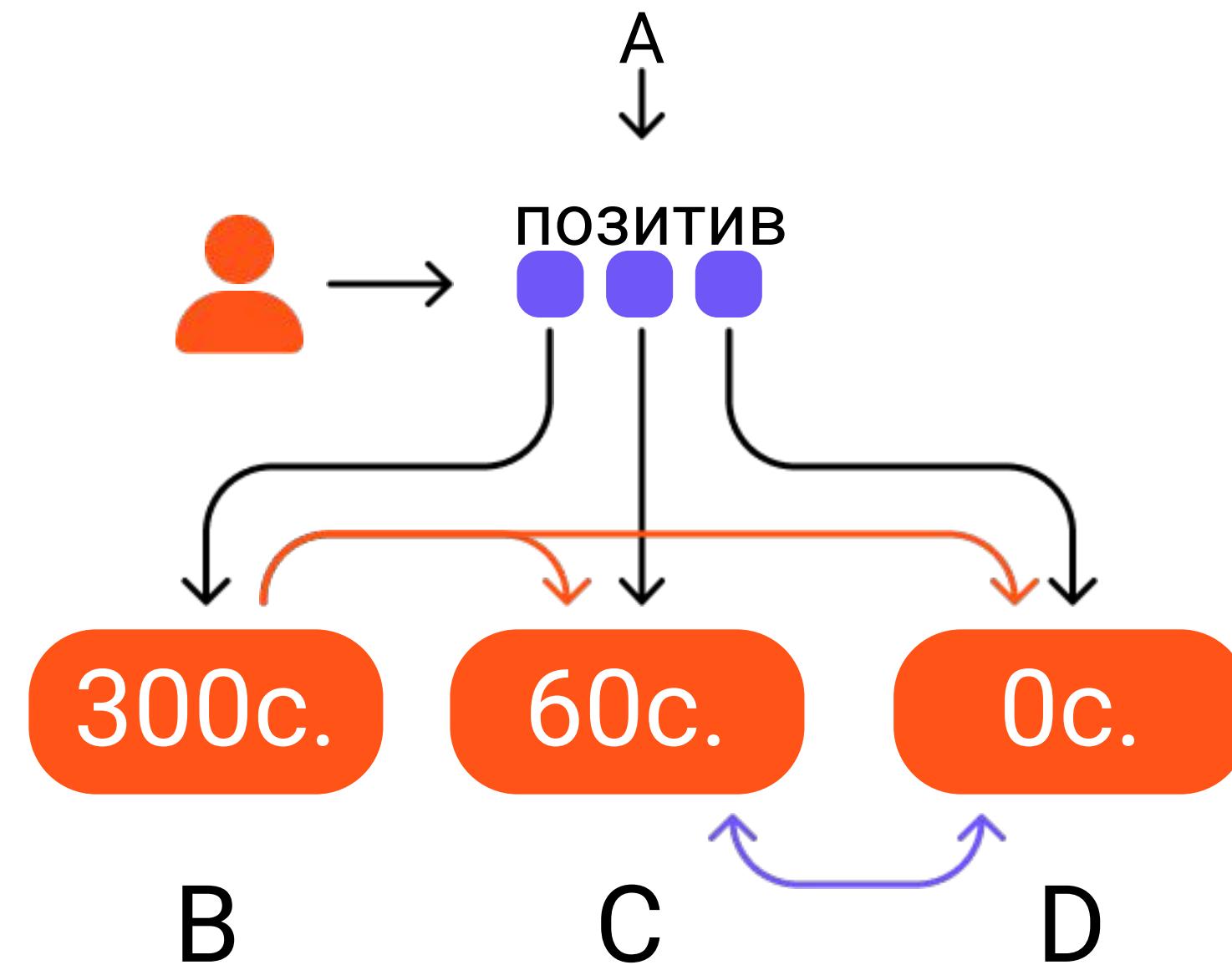


Item2Item: как построить список

Айтемы в списке нужно отранжировать

Решение 2

- ★ залогируем «позитив»

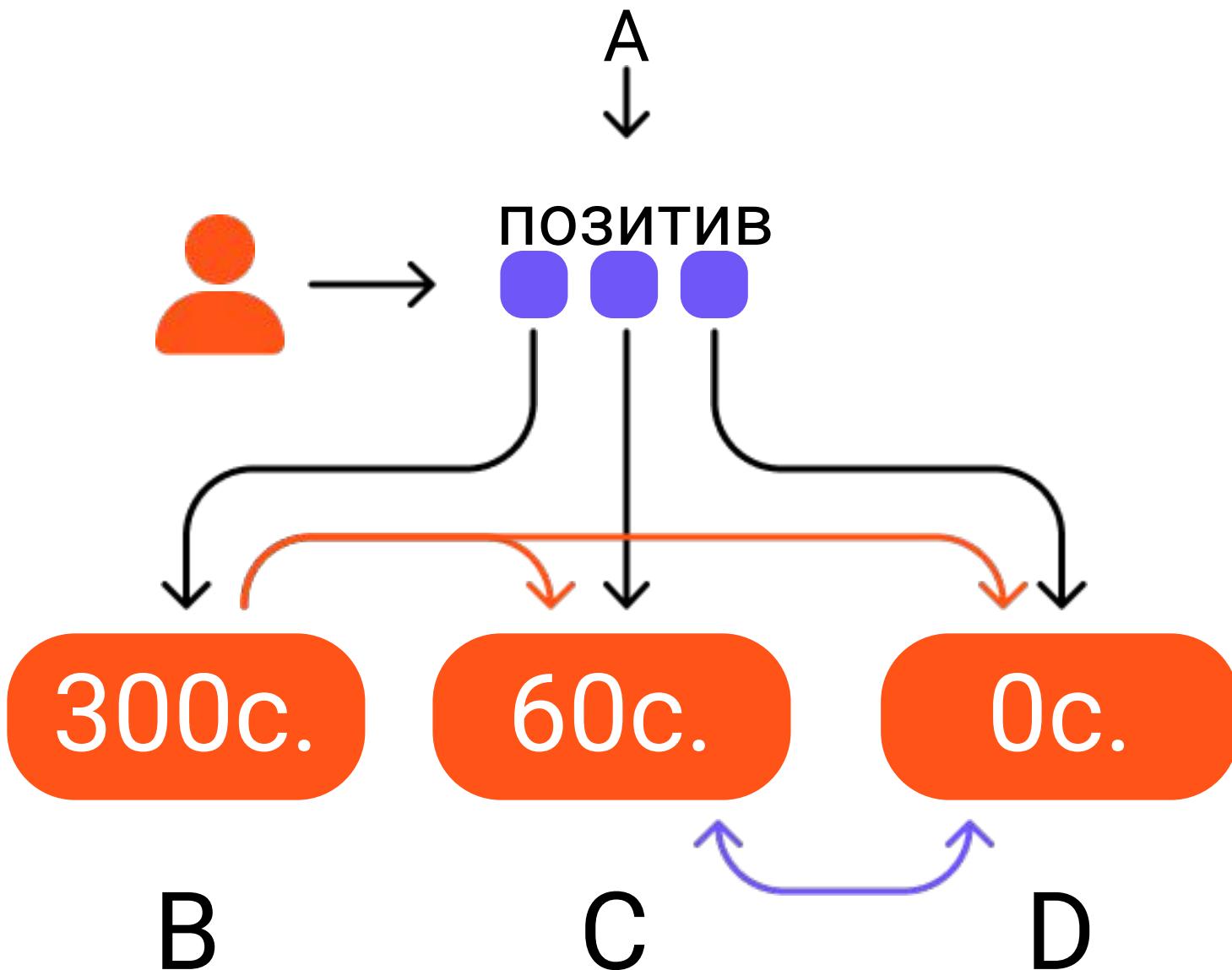


Item2Item: как построить список

Айтемы в списке нужно отранжировать

Решение 2

- ★ залогируем «позитив»
- ★ соберём pairwise датасет
- ★ обучим CatBoost
- ★ фичи
 - ★ коллаборативные похожести
 - ★ контентные похожести
 - ★ авторские похожести



positive	winner	loser	weight
A	B	C	300–60
A	B	D	300
A	C	D	60

Дзен

Saint
HighLoad++
2023



Item2Item: как построить список



Item2Item: как построить список



Item2Item: как построить список

Построенные списки:

- ★ 40 млн позитивов
- ★ 5 млн документов
- ★ максимальная длина 500
- ★ 2 часа
- ★ ~200 Гб



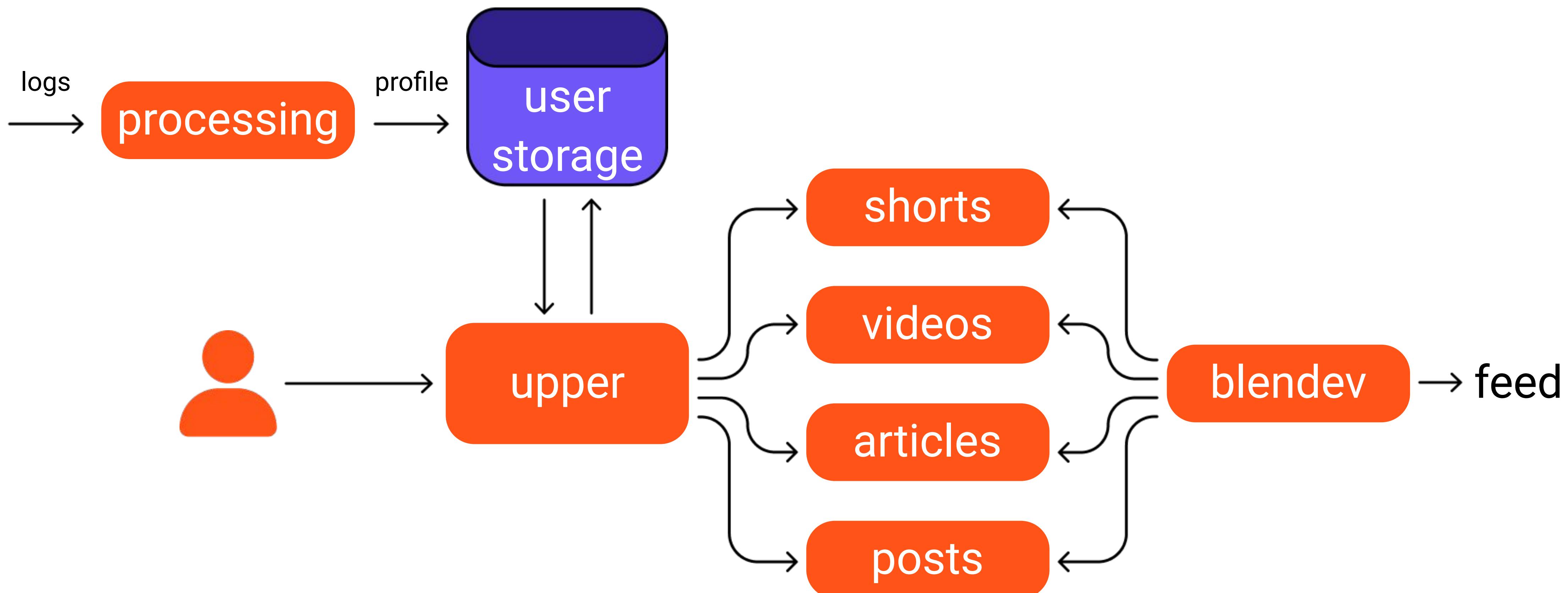
Saint
HighLoad++
2023



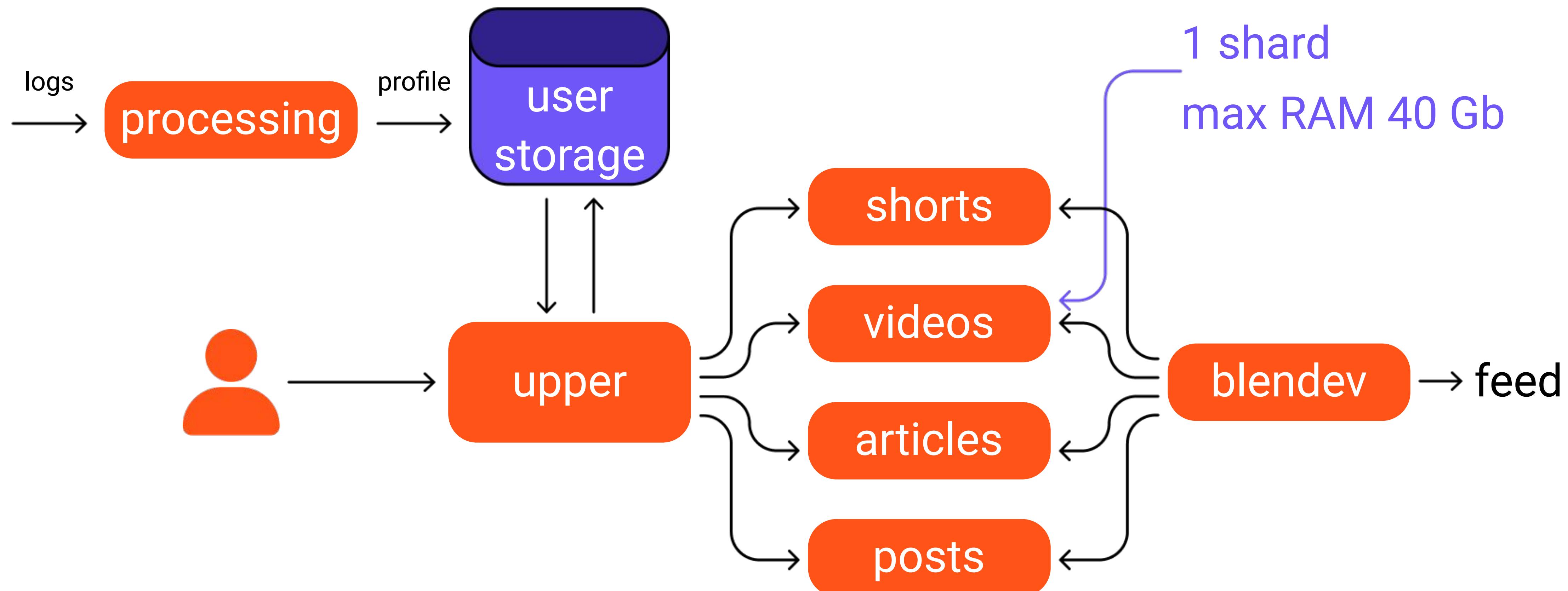
Рекомендер Дзена и и2и- сервис



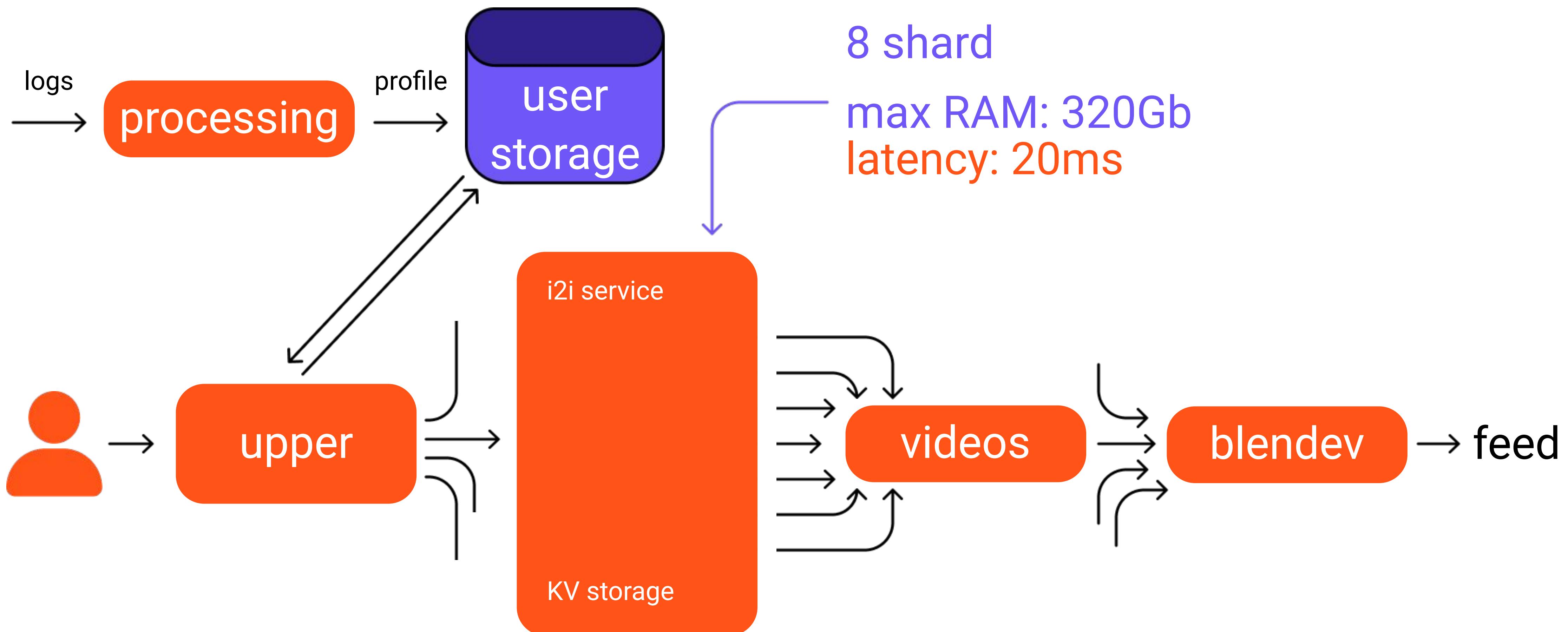
Рантайм рекомендаций в Дзене



Рантайм рекомендаций в Дзене

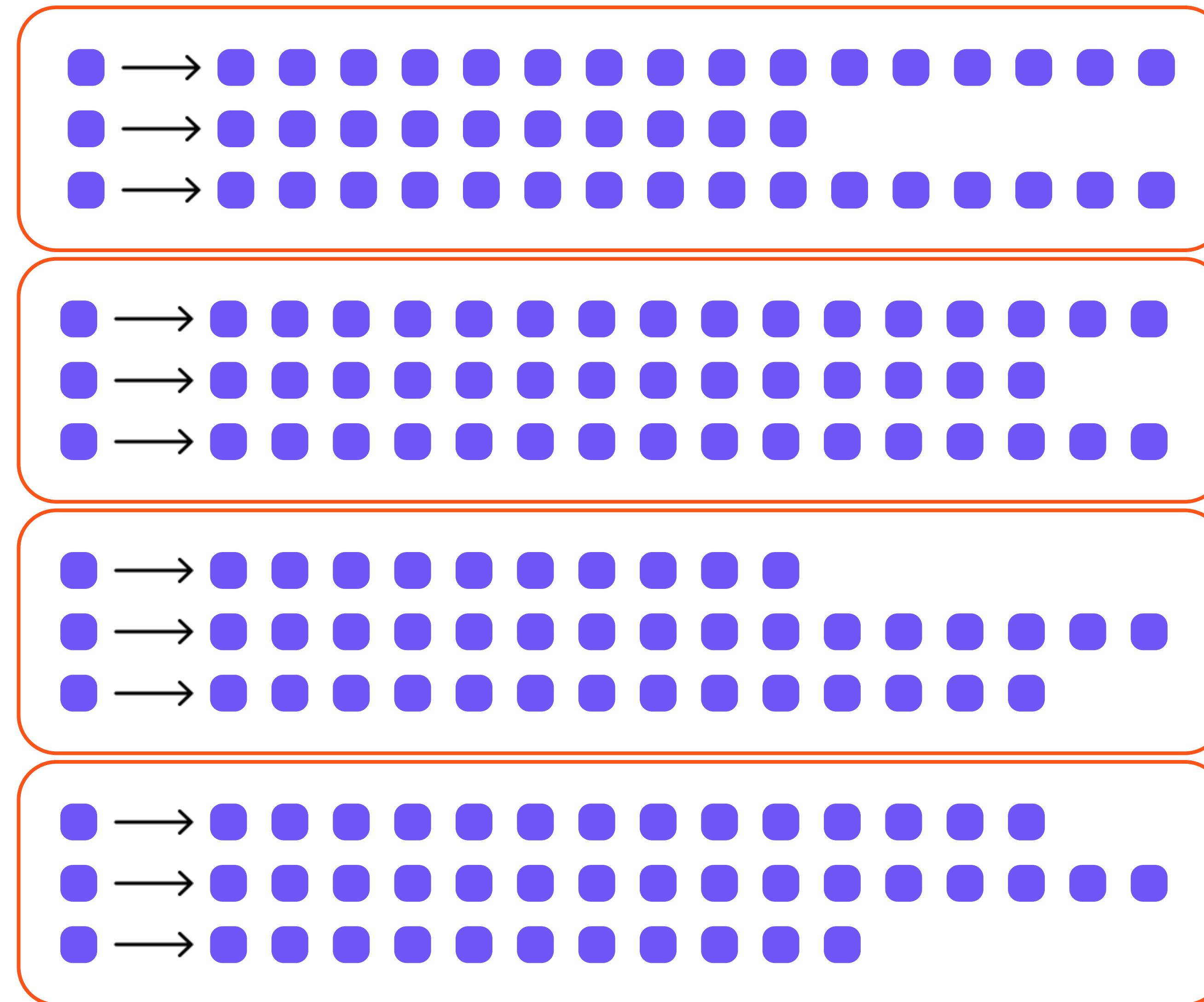


Рантайм рекомендаций в Дзене

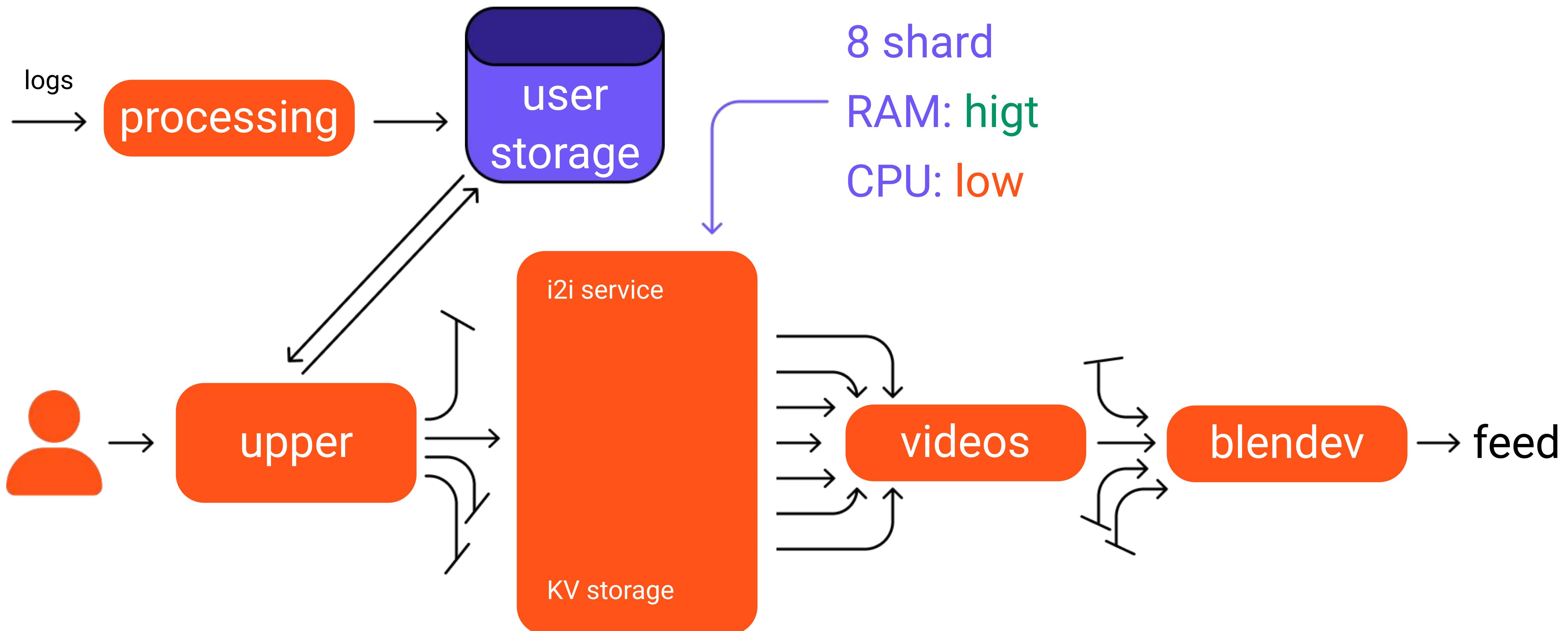


Item2Item сервис

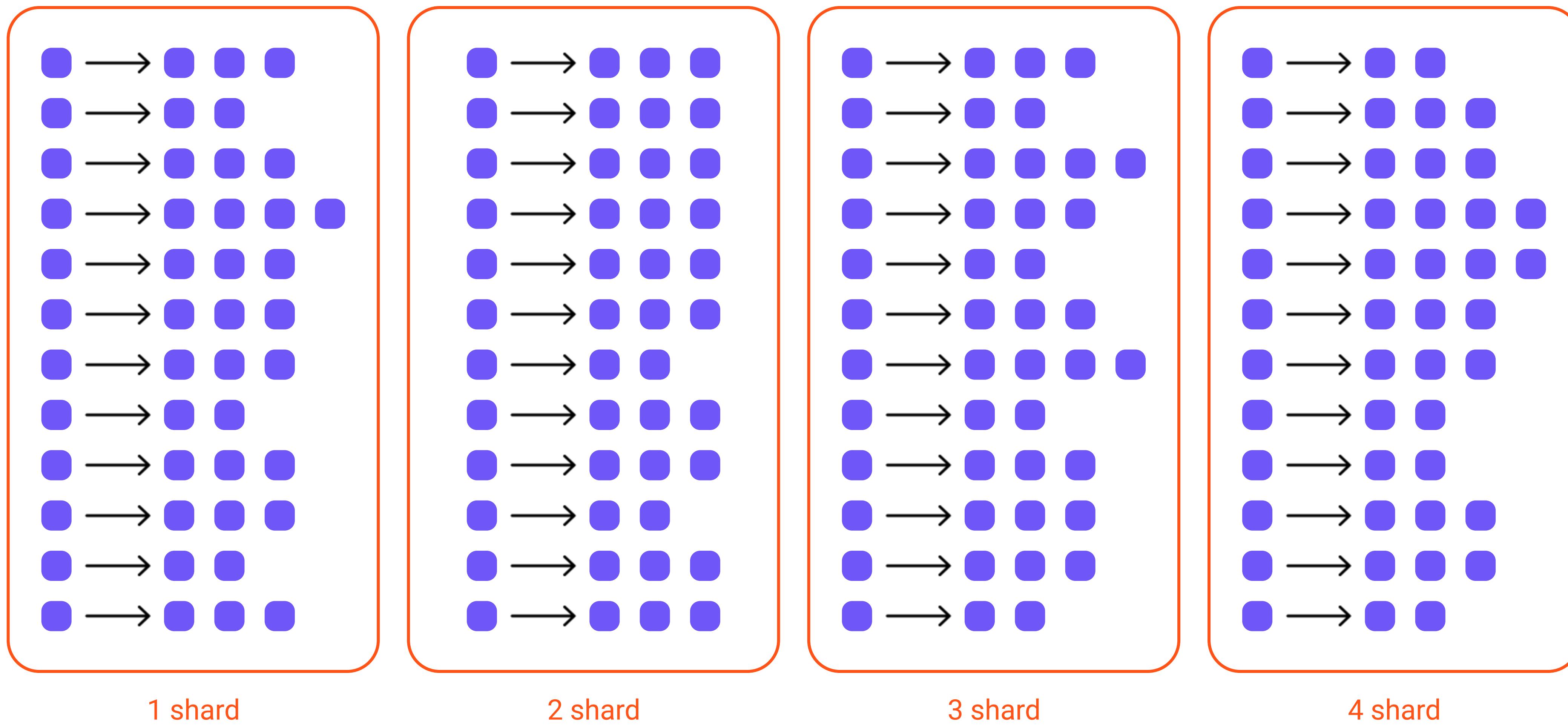
КЛЮЧ →



Item2Item сервис



Item2Item сервис



Saint
HighLoad++
2023



Item2Item сервис

1 шард:

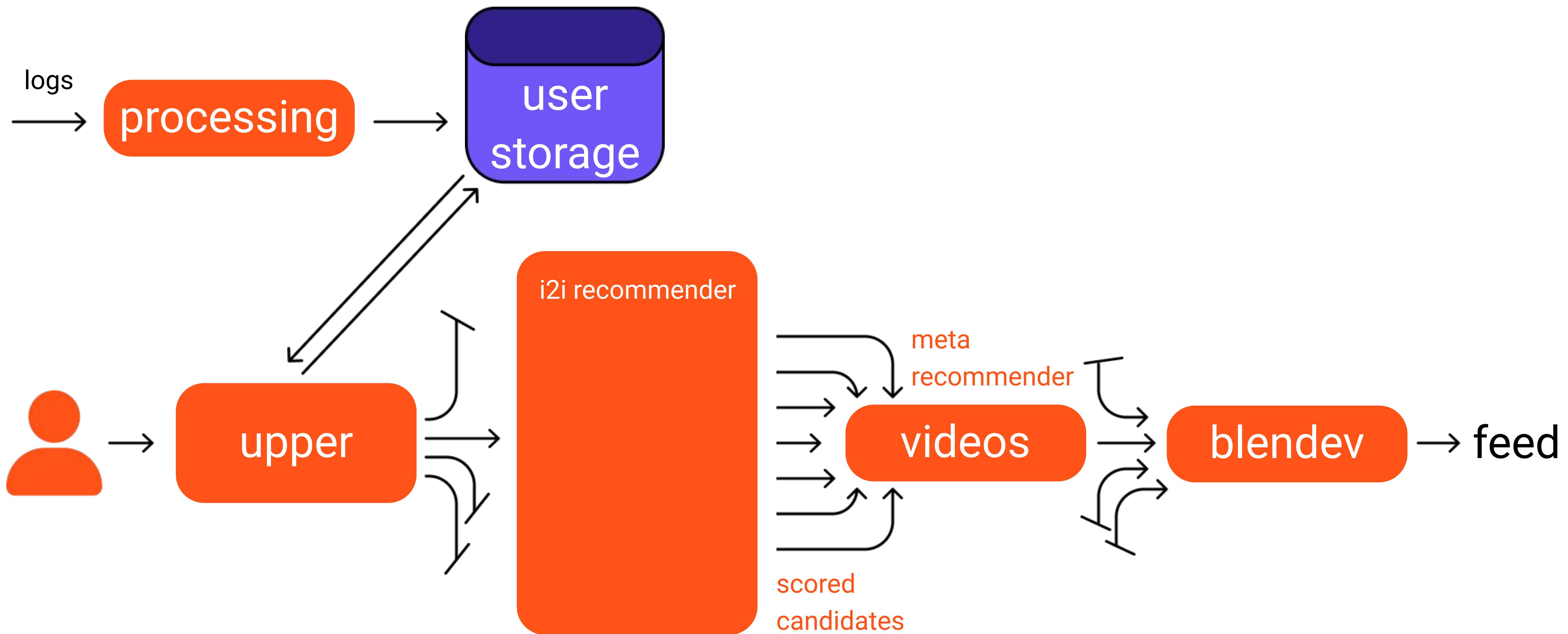
- ★ 1/8 документов
- ★ а если добавить эмбеды?



Saint
HighLoad++
2023



Item2Item рекомендер



Item2Item рекомендер

- ★ легко шардируется
- ★ скорим x8 айтемов
- ★ время ответа рекомендаций даже уменьшилось!
- ★ потребовалось дополнительное железо



Saint
HighLoad++
2023



Результаты



Результаты

timespent

+ 5%

похожесть
на историю
пользователя

90%

↑
50%

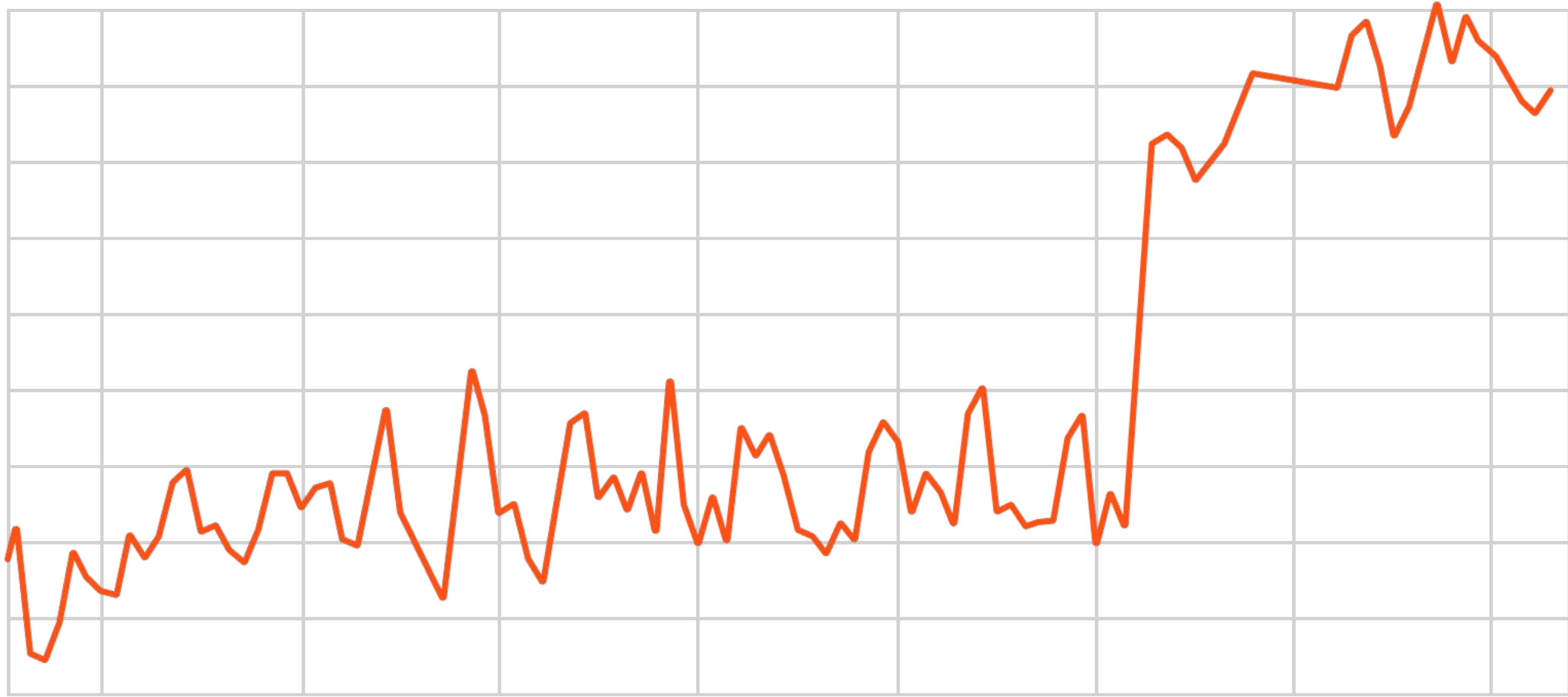


Saint
HighLoad++
2023



NPS

Средняя оценка рекомендаций



Дзен

Saint
HighLoad++
2023



Результаты

Рекомендер стал проще и понятнее

- ★ проще дебажить
- ★ больше телеметрии

12 HNSW

1 i2i list

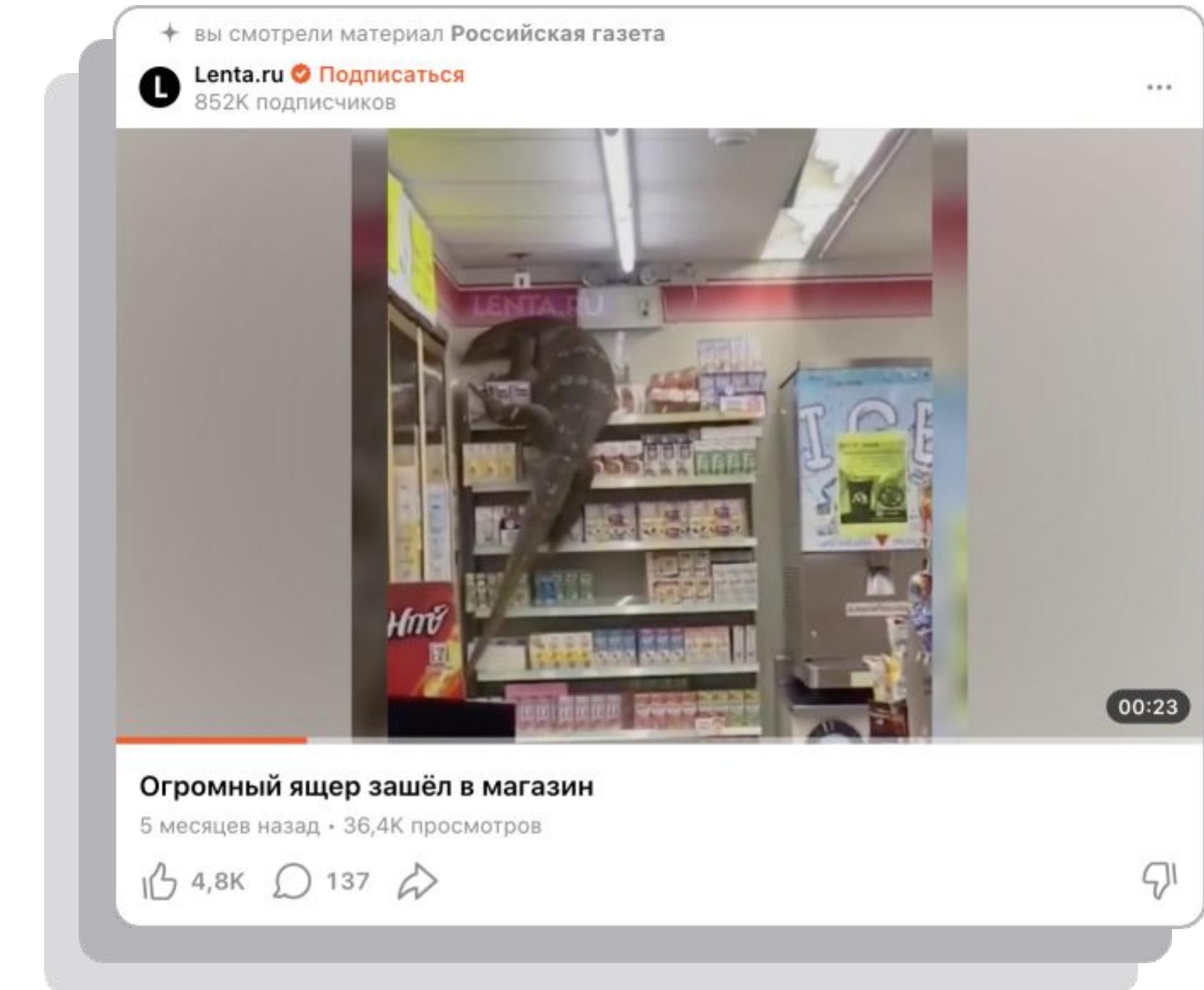


Saint
HighLoad++
2023



Результаты

- ★ рекомендер стал проще и понятнее
- ★ умеем объяснять рекомендации



Дзен

Saint
HighLoad++
2023



Результаты

- ★ рекомендер стал проще и понятнее
- ★ умеем объяснять рекомендации
- ★ при кандидатогенерации гарантируем
использование всей истории



Saint
HighLoad++
2023



Результаты

- ★ рекомендер стал проще и понятнее
- ★ умеем объяснять рекомендации
- ★ при кандидатогенерации гарантируем
использование всей истории
- ★ контролируемый прогрев: баланс между explore и exploit



Saint
HighLoad++
2023

