Яндекс

Яндекс

Краудсорсинг: Как приручить толпу?

Артём Григорьев

Руководитель службы разработки

О докладчике

О докладчике

В петербургском Яндексе с 2010 года

О докладчике

- В петербургском Яндексе с 2010 года
- Разрабатывал инструменты для оценки качества поиска

О докладчике

- В петербургском Яндексе с 2010 года
- Разрабатывал инструменты для оценки качества поиска
- Руководитель службы разработки системы сбора экспертных оценок

О докладчике

- В петербургском Яндексе с 2010 года
- Разрабатывал инструменты для оценки качества поиска
- Руководитель службы разработки системы сбора экспертных оценок
  - Краудсорсинговая платформа Яндекс.Толока

О докладчике

В петербургском Яндексе с 2010 года

Разрабатывал инструменты для оценки качества поиска

Руководитель службы разработки системы сбора экспертных оценок

- Краудсорсинговая платформа Яндекс.Толока
- Сервисы для асессоров, контент-менеджеров, поддержки пользователей

О докладчике

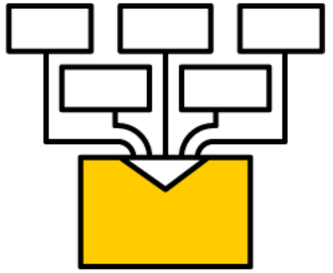
В петербургском Яндексе с 2010 года

Разрабатывал инструменты для оценки качества поиска

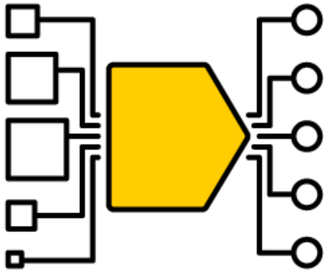
Руководитель службы разработки системы сбора экспертных оценок

- Краудсорсинговая платформа Яндекс.Толока
- Сервисы для асессоров, контент-менеджеров, поддержки пользователей
- Инфраструктура для сбора данных

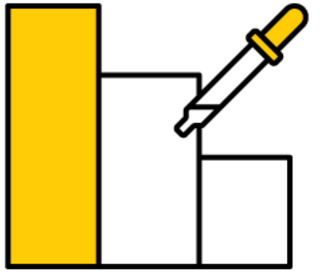
Ручной труд



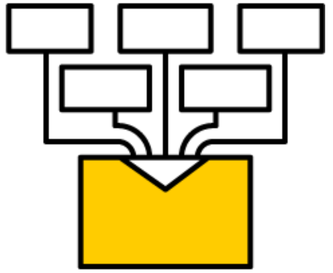
Сбор данных



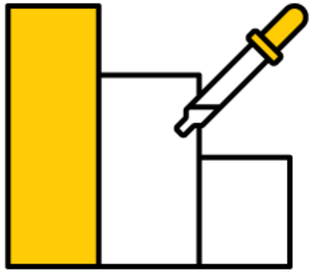
Обучение



Измерение



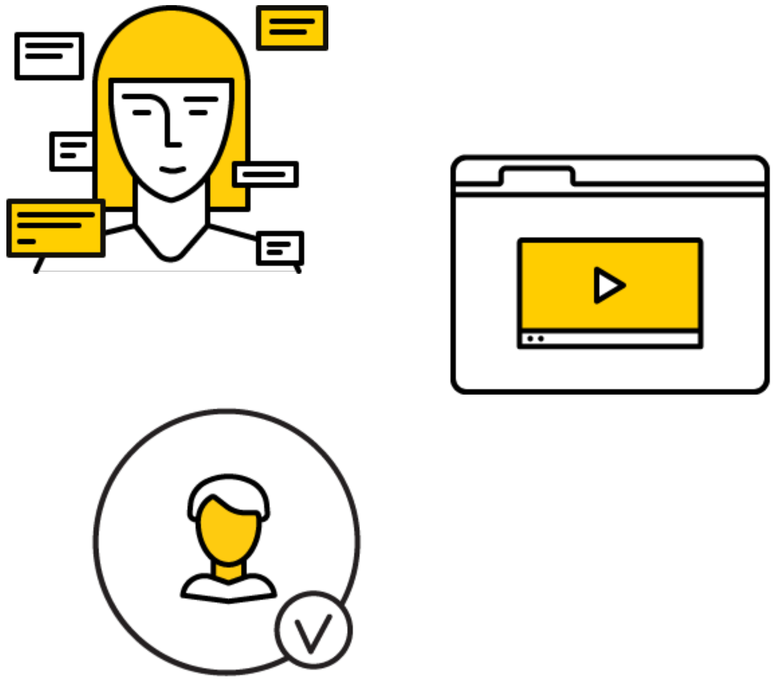
Сбор данных



Измерение

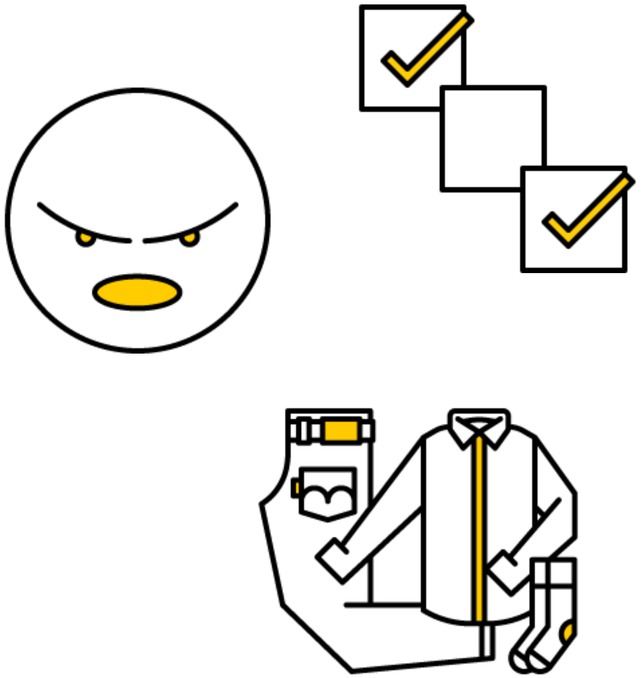
Модерация контента

- Комментарии, посты
- Фотографии, видеозаписи
- Профили, аккаунты



Категоризация

- Изображения
- Эмоциональная окраска текста
- Товары в Интернет-магазинах

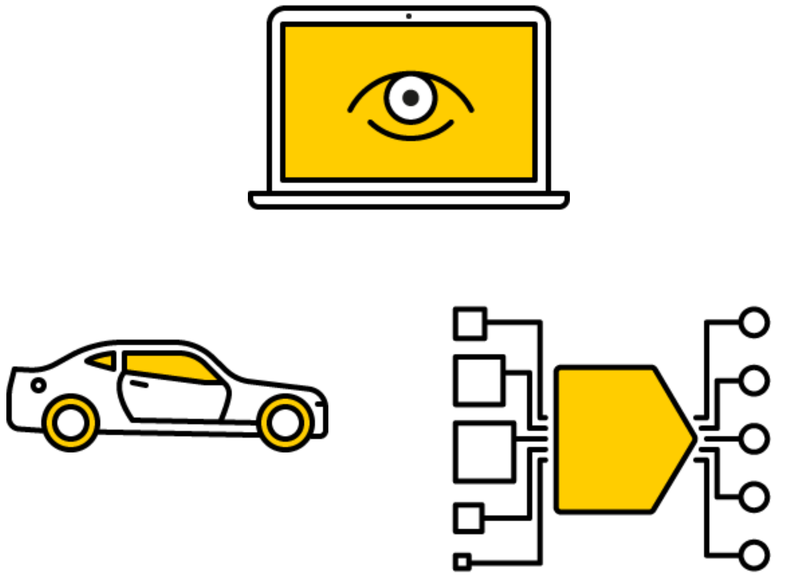


Сбор и обработка данных

- Поиск информации в Интернете
- Описание и атрибуты товаров
- Валидация данных

Аннотирование изображений

- Компьютерное зрение
- Самоуправляемые автомобили



Транскрипция

- Перевод аудиозаписей в текст
- Расшифровка надписей на изображениях

Ручной труд

Ручной труд

- Долго

Ручной труд

- Долго
- Дорого

Ручной труд

- Долго
- Дорого
- Бюрократия

Краудсорсинг

1

2

Краудсорсинг

Способ привлечь к решению различных задач неограниченную группу неизвестных исполнителей

Волонтёрский краудсорсинг

Backyard Worlds: Planet

https://www.zooniverse.org/projects/marckuchner/backyard-worlds-planet-9

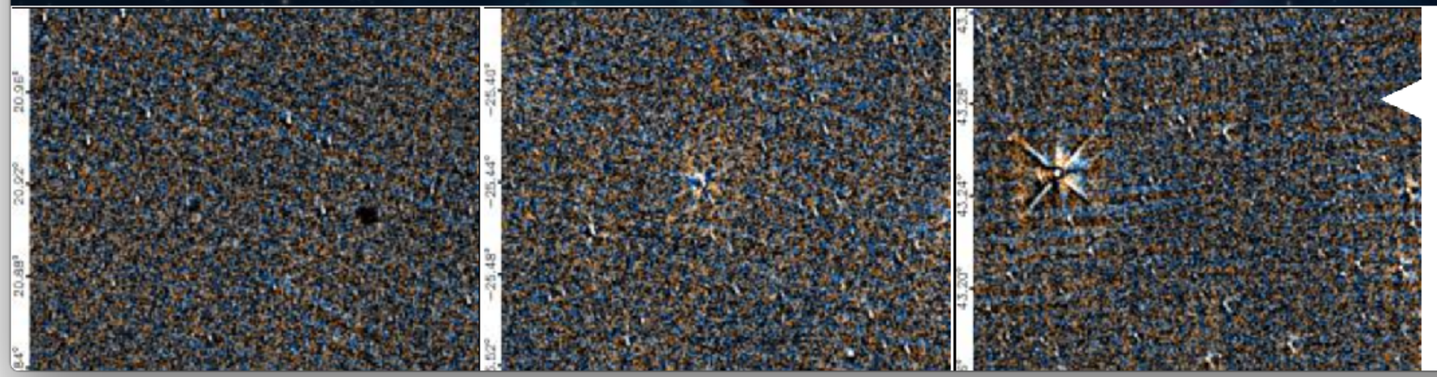
BACKYARD WORLDS: PLANET 9

ABOUT CLASSIFY TALK COLLECT BLOG F.A.Q.

Help scan the realm beyond Neptune for brown dwarfs and planet nine.

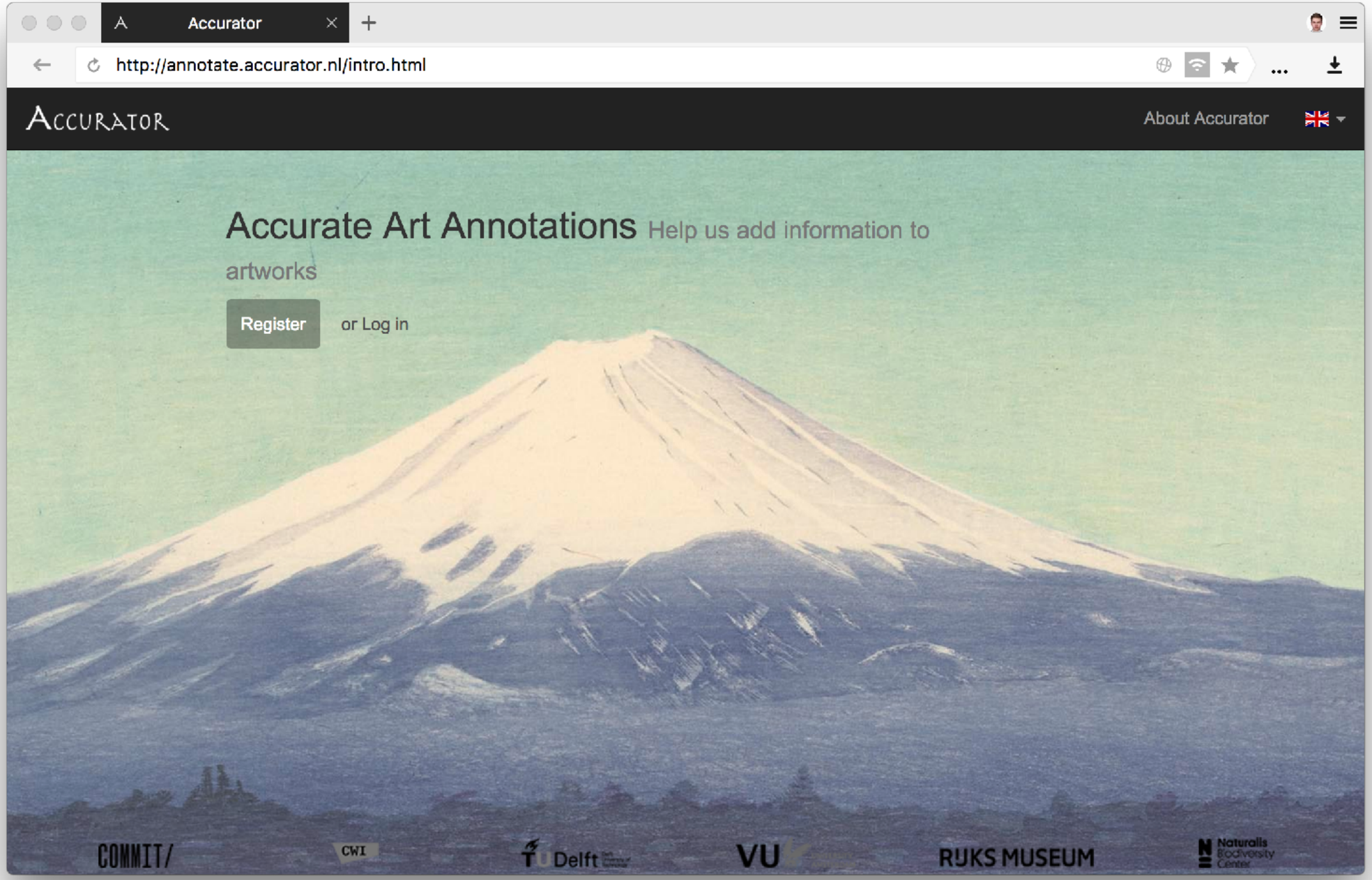
Learn more

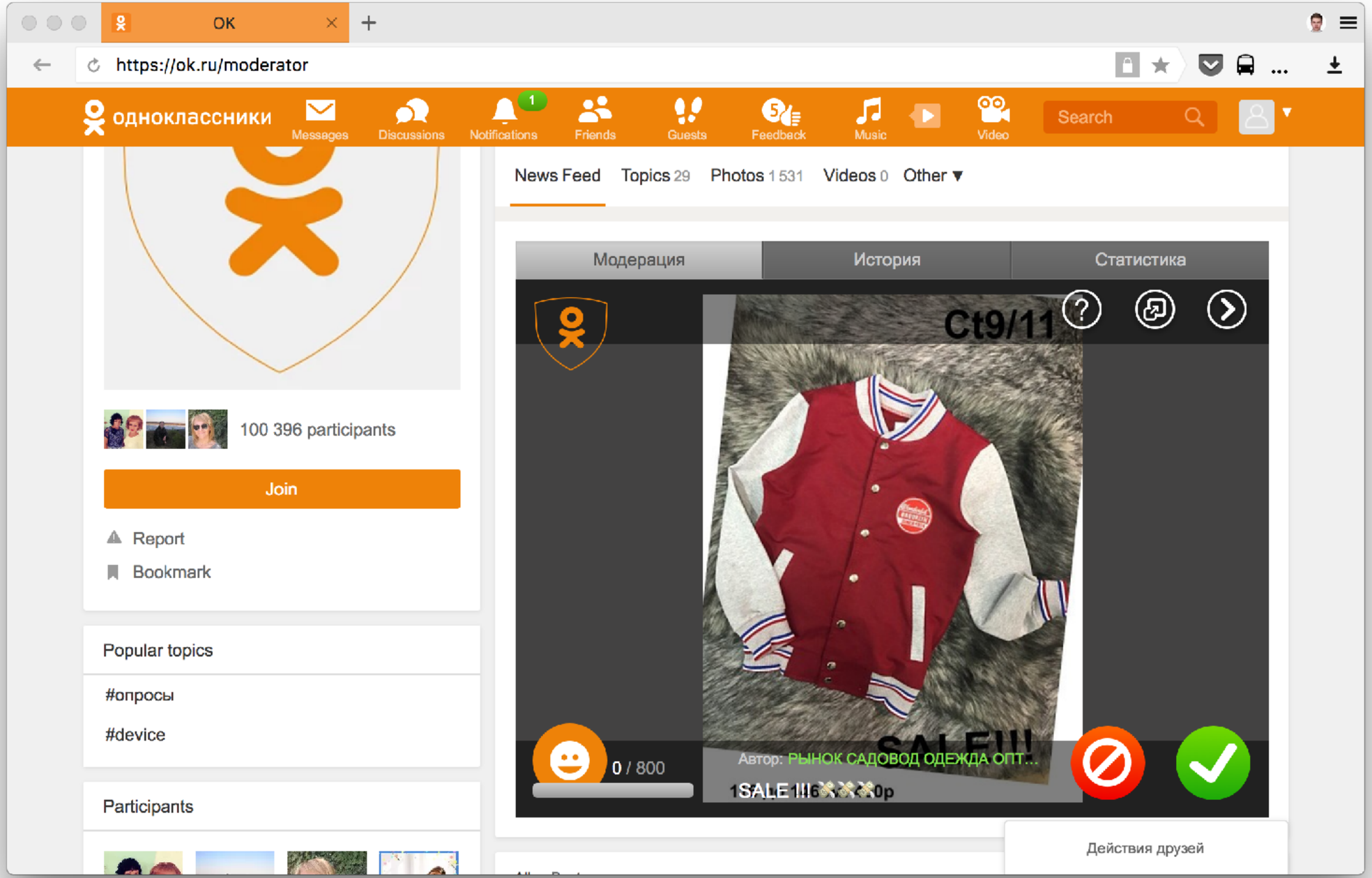
Get started



23 people are talking about **Backyard Worlds: Planet 9** right now.

Join in


Accurator
http://annotate.accurator.nl/intro.html
ACCURATOR
About Accurator
Accurate Art Annotations Help us add information to artworks
Register
or Log in
COMMIT/
CWI
Delft
VU
RIJKS MUSEUM
Naturalis Biodiversity Center


OK
https://ok.ru/moderator
одноклассники
Messages
Discussions
Notifications
Friends
Guests
Feedback
Music
Video
Search
News Feed
Topics 29
Photos 1 531
Videos 0
Other
100 396 participants
Join
Report
Bookmark
Popular topics
#опросы
#device
Participants
Модерация
История
Статистика
Ct9/11
0 / 800
Автор: РЫНОК САДОВОД ОДЕЖДА ОПТ...
SALE!!!
Действия друзей

Волонтёрский краудсорсинг

- Социальная мотивация
- Нет проблем с качеством

Волонтёрский краудсорсинг

- Социальная мотивация
- Нет проблем с качеством

- **Но** плохо масштабируется горизонтально

Платный краудсорсинг

amazon
mechanical turk
beta

CrowdFlower

clickworker

spare 5

Mighty Ai

Нет русскоязычных исполнителей!

Яндекс.Толока

Толока

Древний славянский обычай собираться вместе для выполнения большой общественно-значимой работы

Рынок

Исполнитель **Платформа** **Заказчик**

Обработка платежей, поддержка

Размещение заданий

Размещение заданий

- Выбрать или создать интерфейс

Размещение заданий

- Выбрать или создать интерфейс
- Загрузить задания

Размещение заданий

- Выбрать или создать интерфейс
- Загрузить задания
- Задать настройки, фильтры и контроль качества

Размещение заданий

- Выбрать или создать интерфейс
- Загрузить задания
- Задать настройки, фильтры и контроль качества
- Запустить и дождаться выполнения

Размещение заданий

- Выбрать или создать интерфейс
- Загрузить задания
- Задать настройки, фильтры и контроль качества
- Запустить и дождаться выполнения
- Скачать результаты

Примеры заданий

Задания

Классификация

1 Гео 2 не Гео

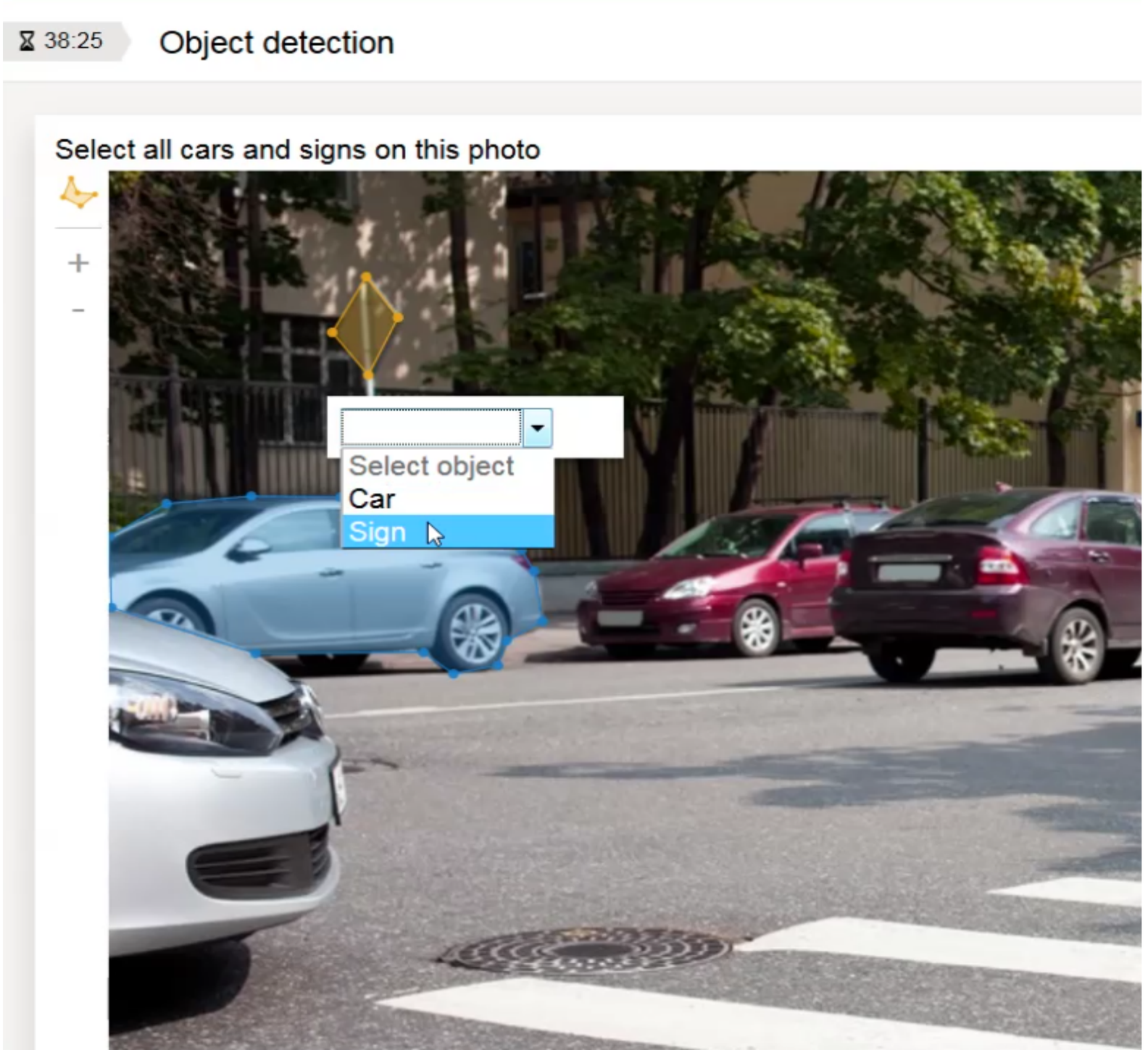
1 Гео 2 не Гео

1 Гео 2 не Гео

1 Гео 2 не Гео

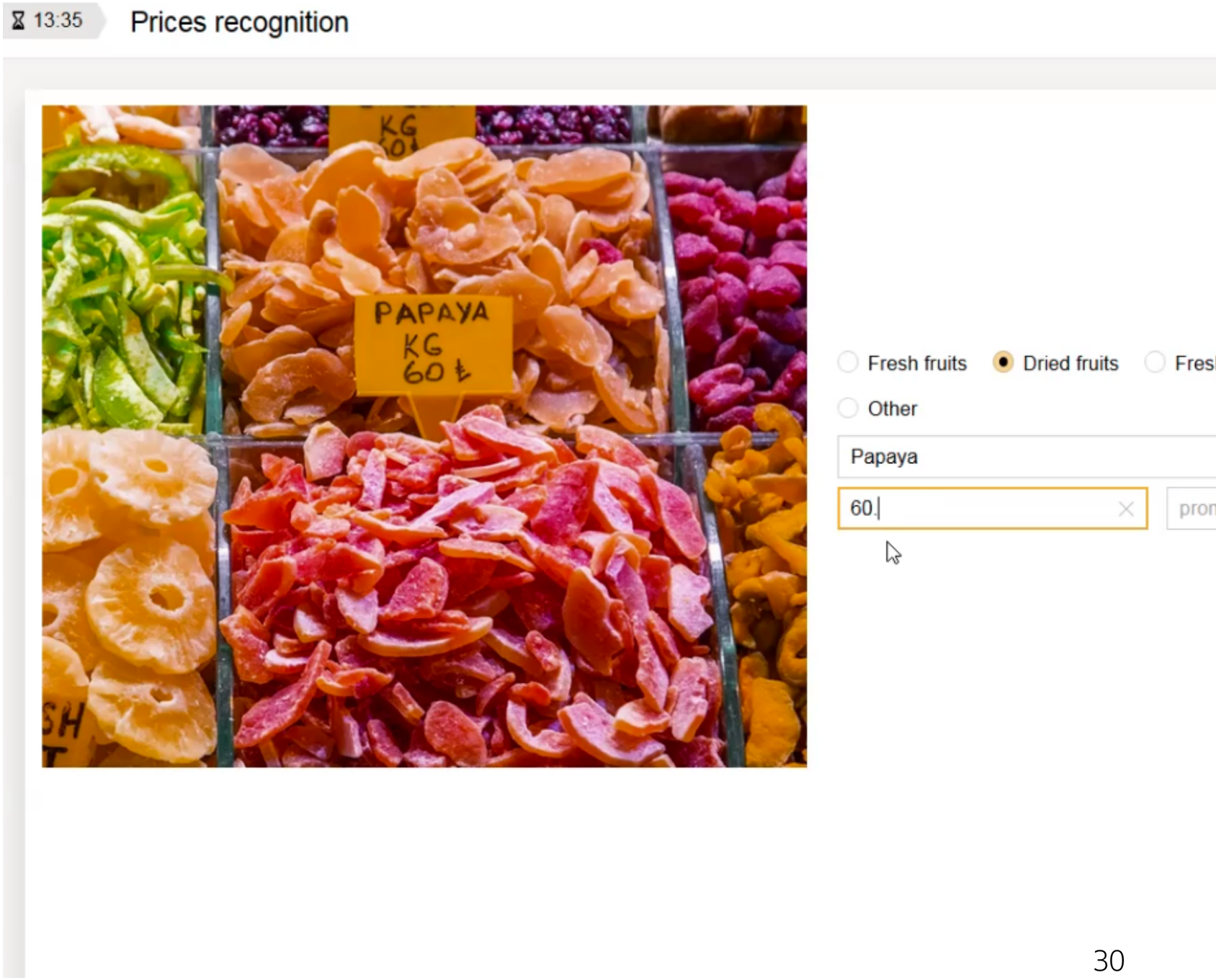
Задания

Выделение объектов



Задания

Распознавание текстов



Поиграем в толокеров?

Основано на реальных заданиях

Какая это одежда?

1. Футболка
2. Рубашка с коротким рукавом
3. Туника
4. Футболка-поло

Какая это одежда?

1. Футболка
2. Рубашка с коротким рукавом
3. Туника
4. **Футболка-поло**

Какая это обувь?

1. Челси
2. Броги
3. Лоферы
4. Монки

Какая это обувь?

1. Челси
2. **Броги**
3. Лоферы
4. Монки

Какое это платье?

1. Повседневное
2. Деловое
3. Коктейльное
4. Клубное

Какое это платье?

1. Повседневное
2. Деловое
3. **Коктейльное**
4. Клубное

Выбор исполнителей

Выбор исполнителей

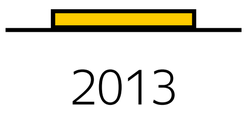
- Тренировка и экзамен, по результатам которого исполнителю назначается **навык**

Выбор исполнителей

- Тренировка и экзамен, по результатам которого исполнителю назначается **навык**
- Доступ к заданиям можно ограничить при помощи фильтров

Главный плюс краудсорсинга

Задания



Проекты

Исполнители

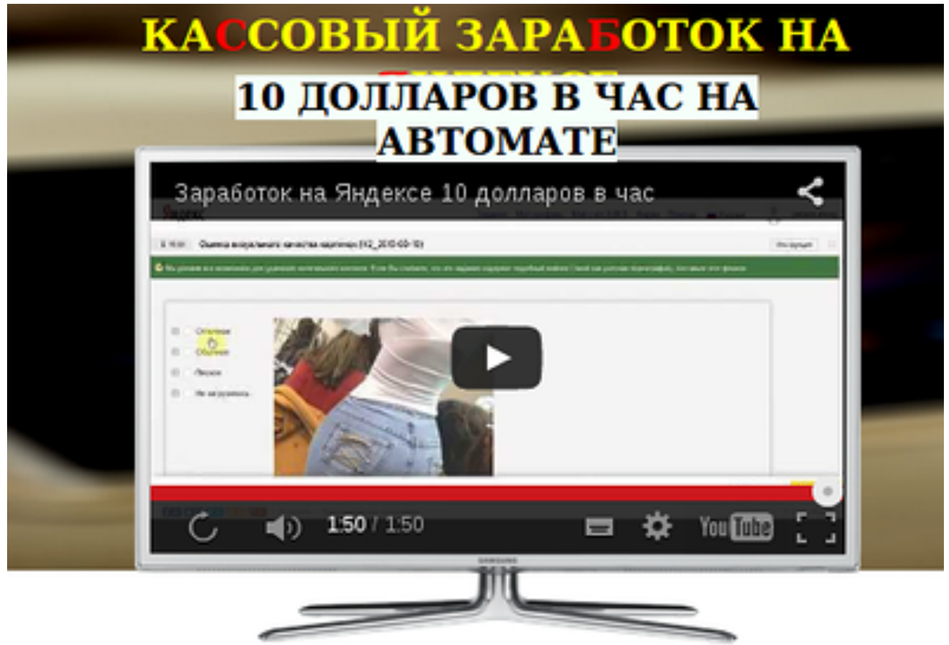
Бесконечное масштабирование!

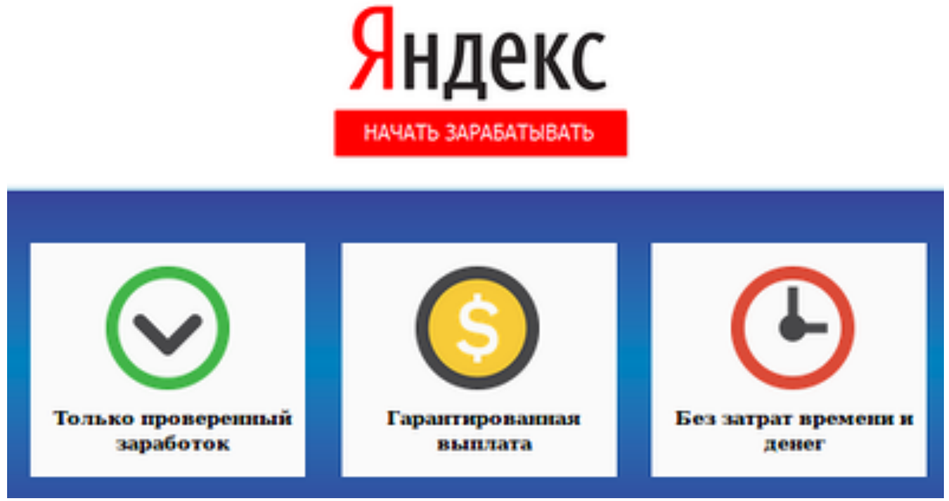
Большое спам-давление :(

- Читеры
- Неграмотные
- Невнимательные
- Боты

Большое спам-давление :(

- Читеры
- Неграмотные
- Невнимательные
- Боты

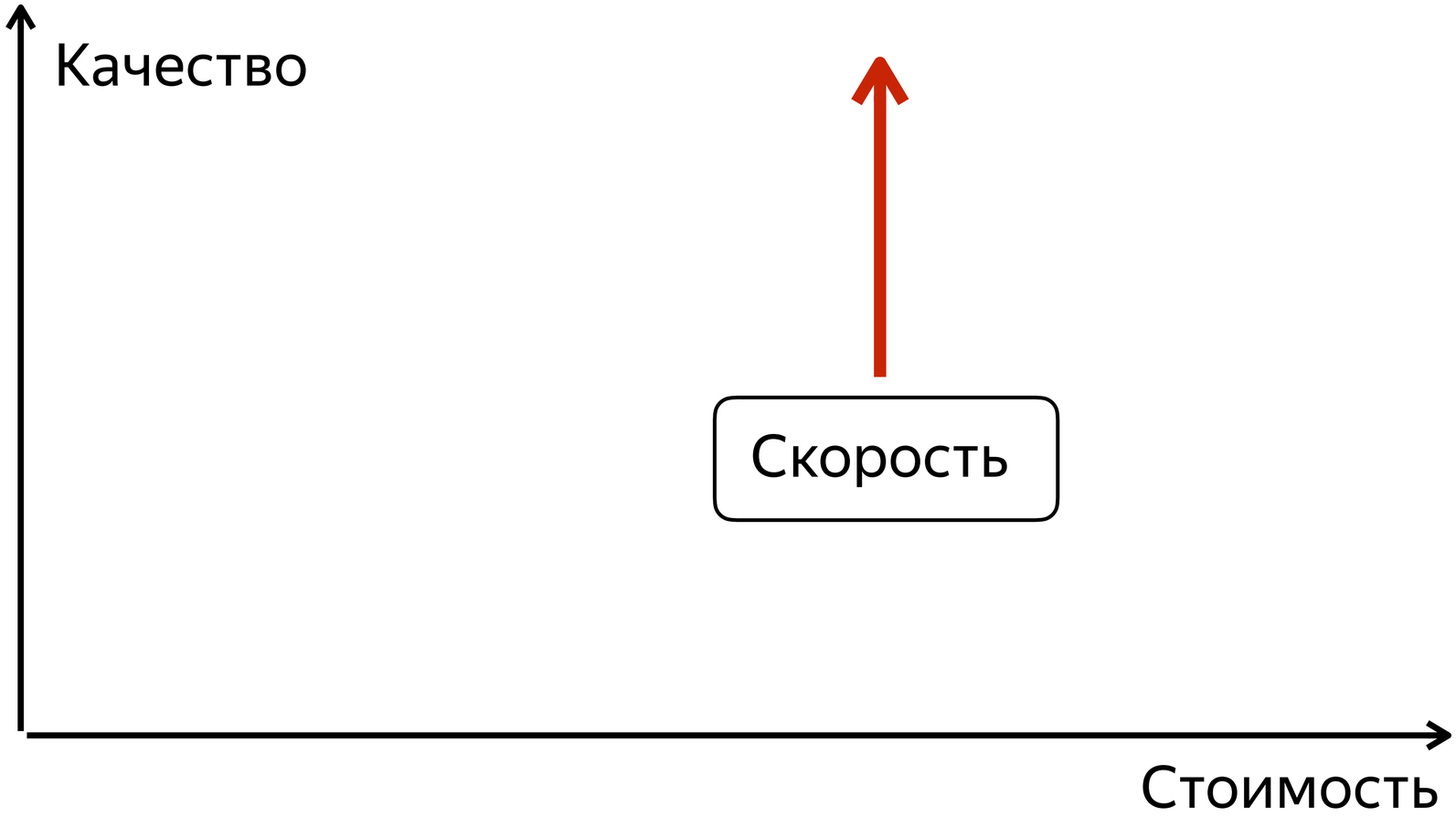




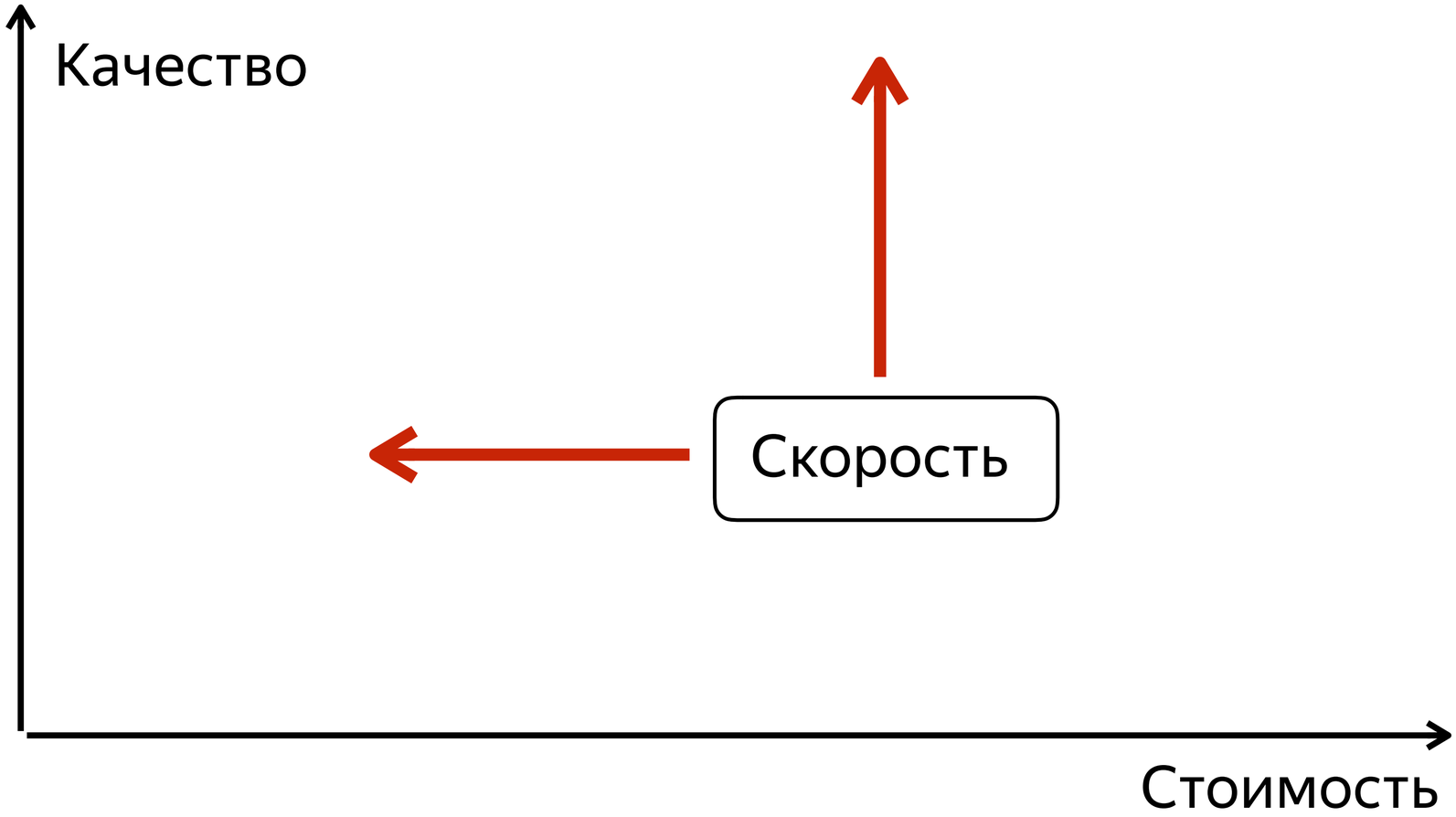
Что делать?

Что делать?

Что делать?



Что делать?



Качество

Как измерять качество?

Ханипоты

- Honeypot — задание-приманка
- Мы заранее знаем правильный ответ и ищем тех, кто отвечает по-другому

Ханипоты

- Honeypot — задание-приманка
- Мы заранее знаем правильный ответ и ищем тех, кто отвечает по-другому

- Минус: должны быть простыми

Согласованность

- Выдаём задания в «перекрытии»: одно задание нескольким исполнителям
- Сравниваем ответы, находим тех, кто отвечает не как все

Согласованность

- Выдаём задания в «перекрытии»: одно задание нескольким исполнителям
- Сравниваем ответы, находим тех, кто отвечает не как все
- Минус: возникают проблемы с редкими вариантами ответов

Выборочная проверка

- Частично перепроверяем ответы исполнителей
- Равномерная выборка
- Самый точный способ

Выборочная проверка

- Частично перепроверяем ответы исполнителей
- Равномерная выборка
- Самый точный способ

- Минус: самый дорогой

Выборочная проверка

- Частично перепроверяем ответы исполнителей
- Равномерная выборка
- Самый точный способ

- Минус: самый дорогой

* Дальнейшие выводы в докладе основаны на этом способе оценки качества

Как управлять качеством?

Простые способы

Простые способы

- Captcha

Простые способы

- Captcha
- Эвристики по скорости выполнения заданий

Простые способы

- Captcha
- Эвристики по скорости выполнения заданий
- Эвристики по одинаковым ответам

Способы посложнее

Способы посложнее

- Правильно подобранные ханипоты

Способы посложнее

- Правильно подобранные ханипоты
- Бан при рассогласованности с другими исполнителями

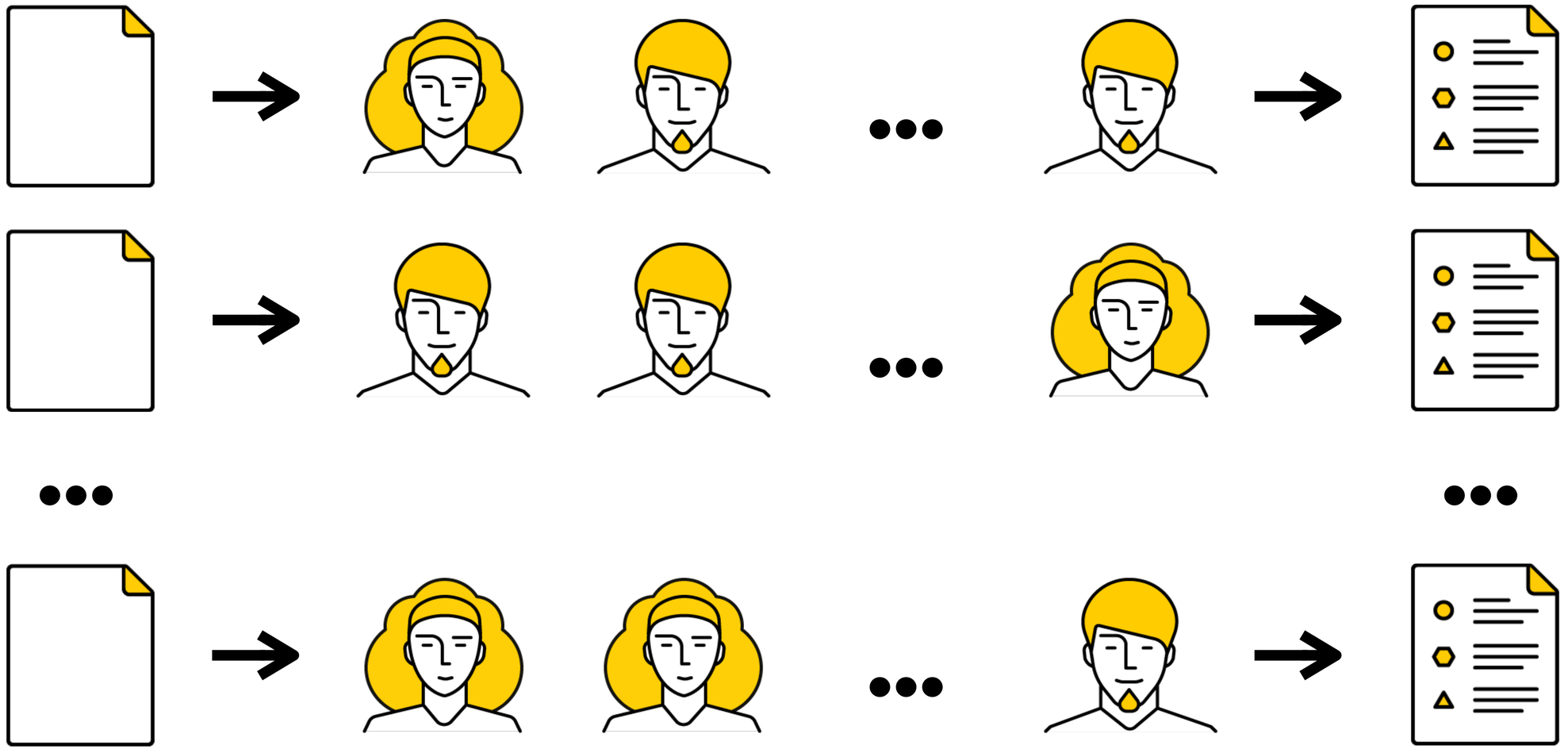
Способы посложнее

- Правильно подобранные ханипоты
- Бан при рассогласованности с другими исполнителями
- Взвешенная метрика точности

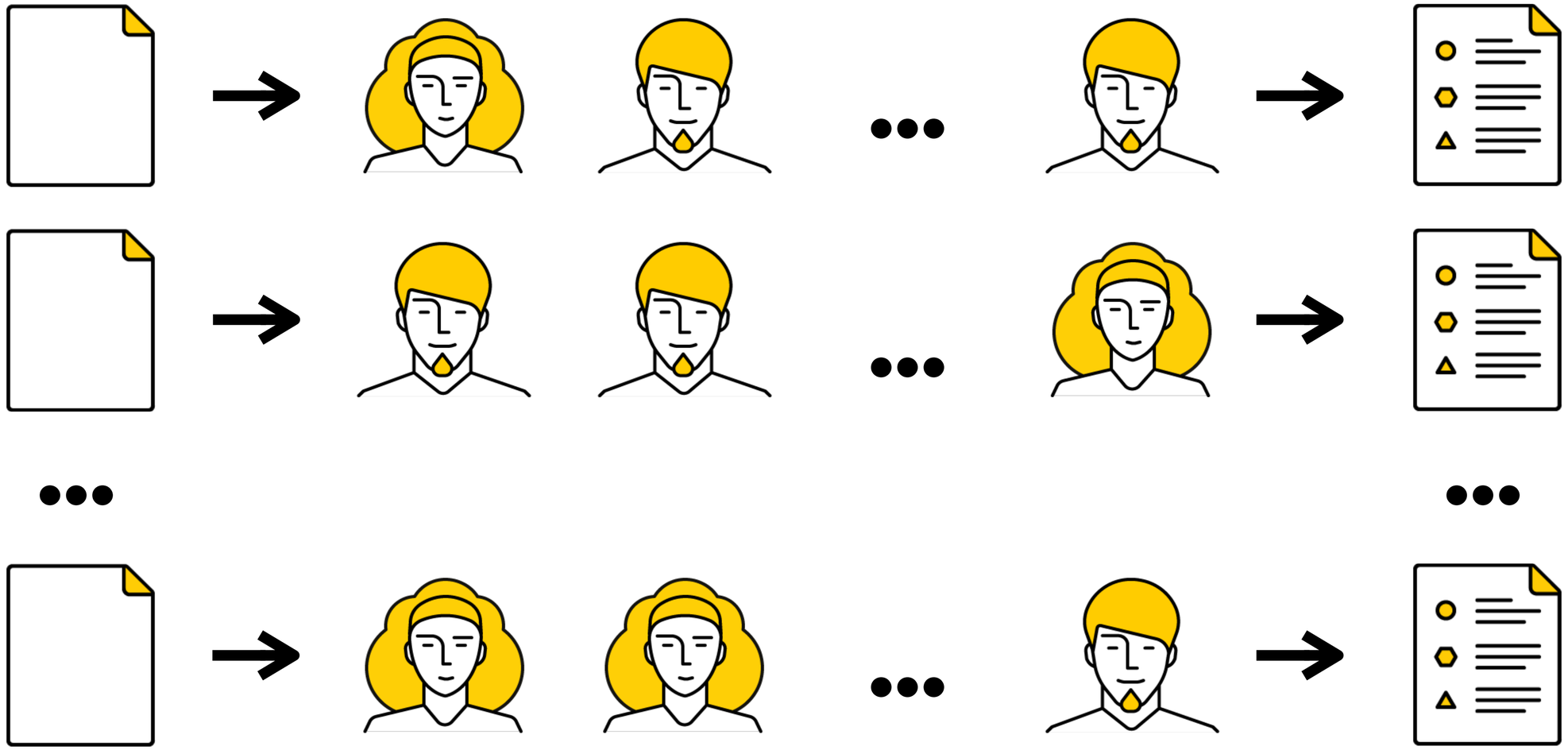
Умные модели агрегации ответов

Улучшаем качество данных без увеличения затрат

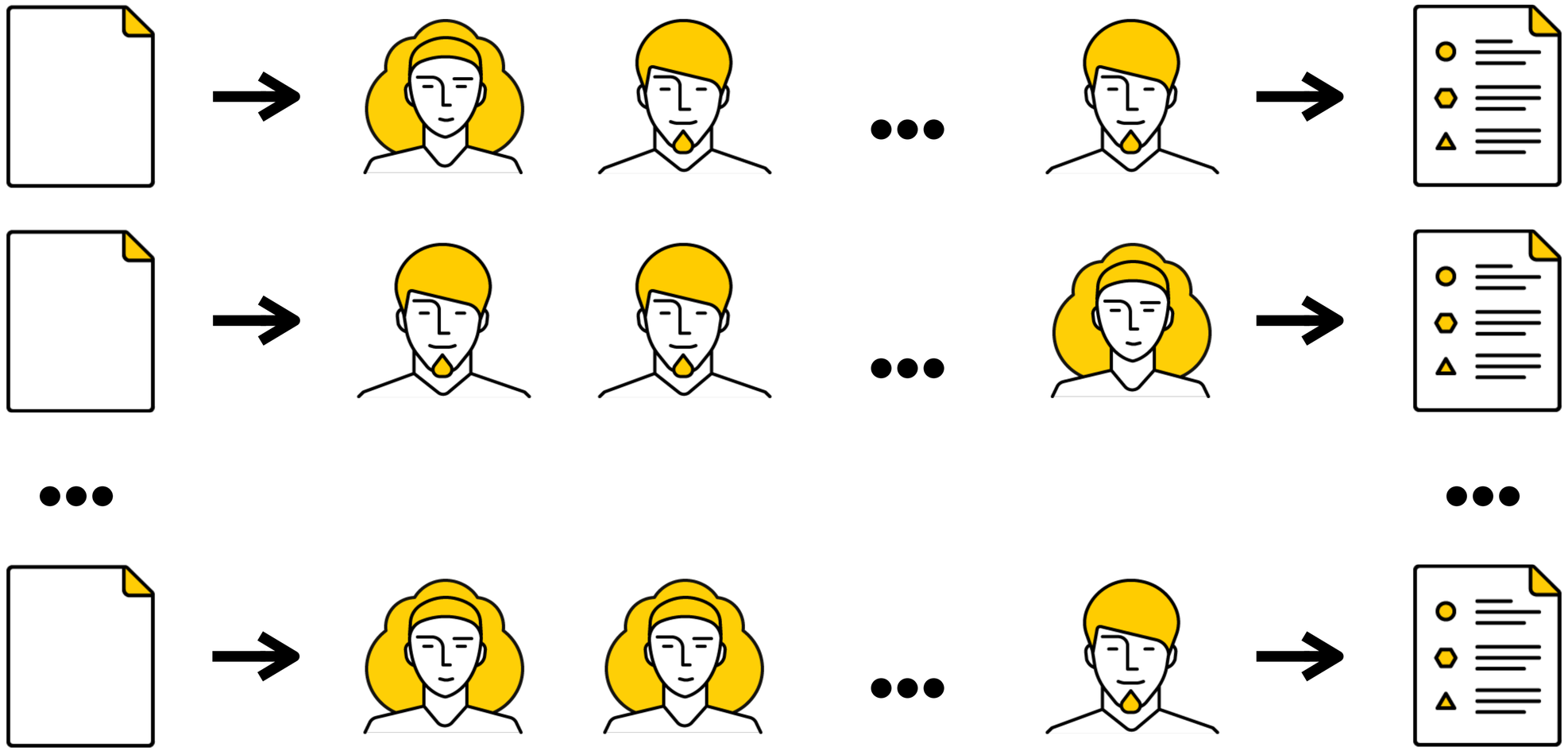
Агрегация ответов



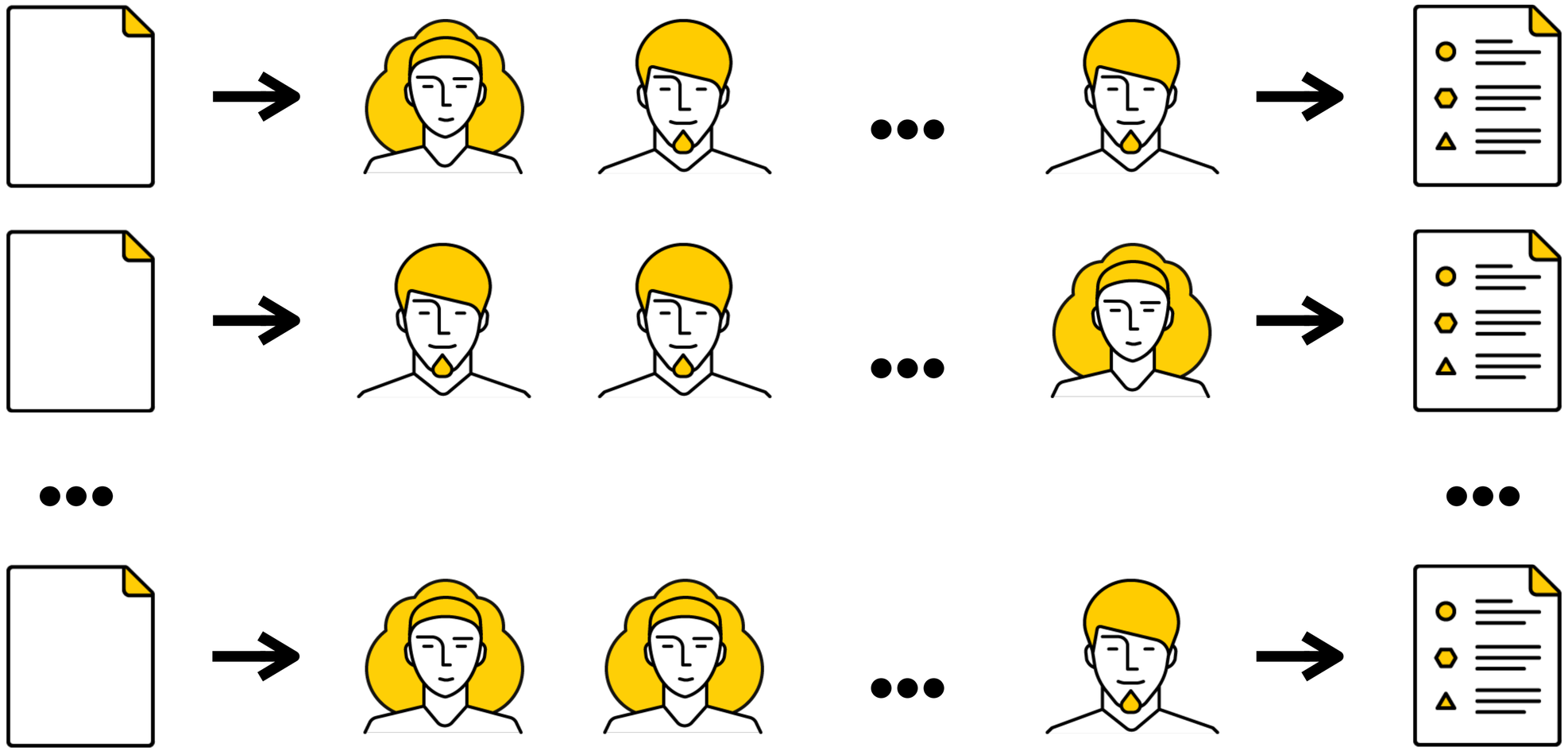
P, d_j L_{wj}, e_w



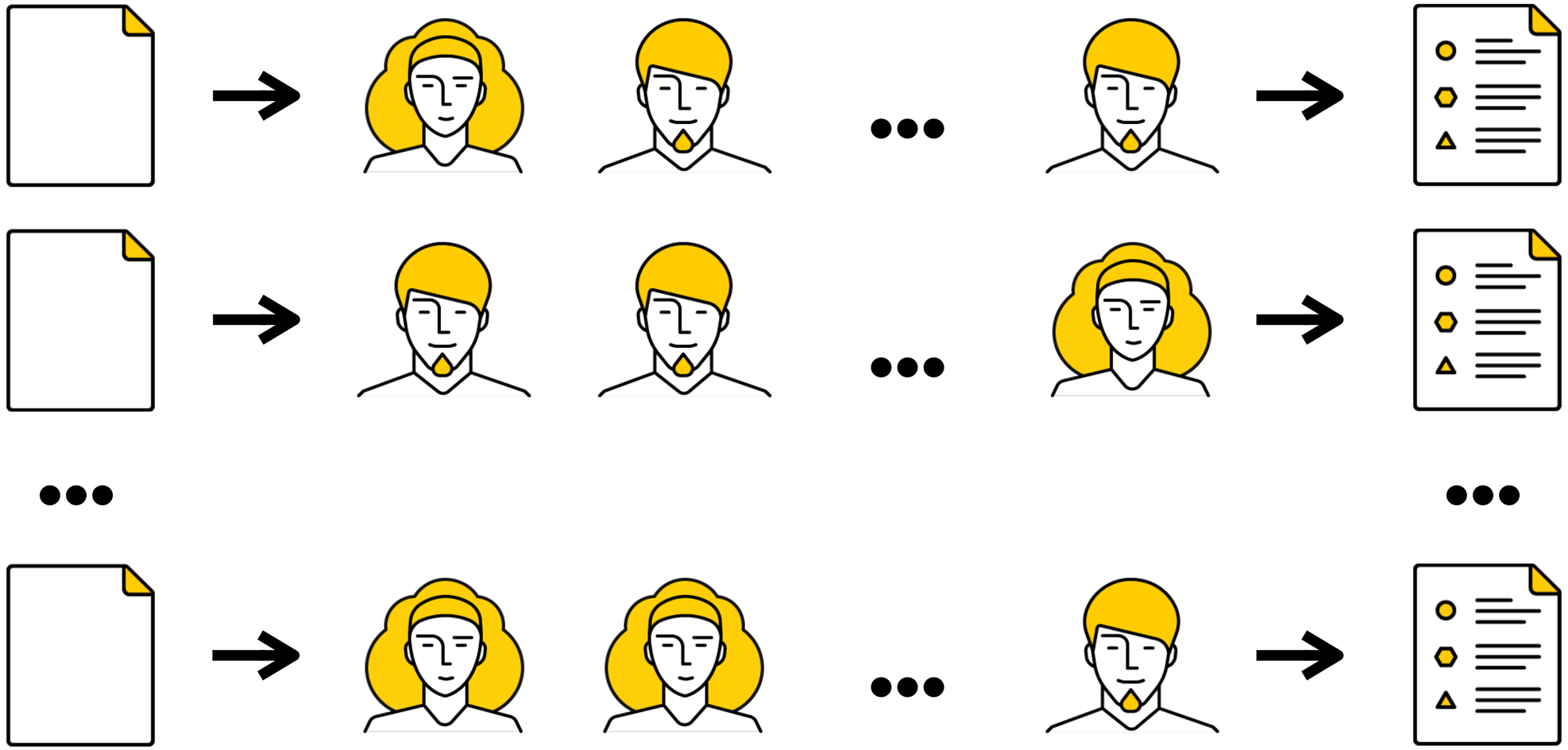
P, d_j L_{wj}, e_w



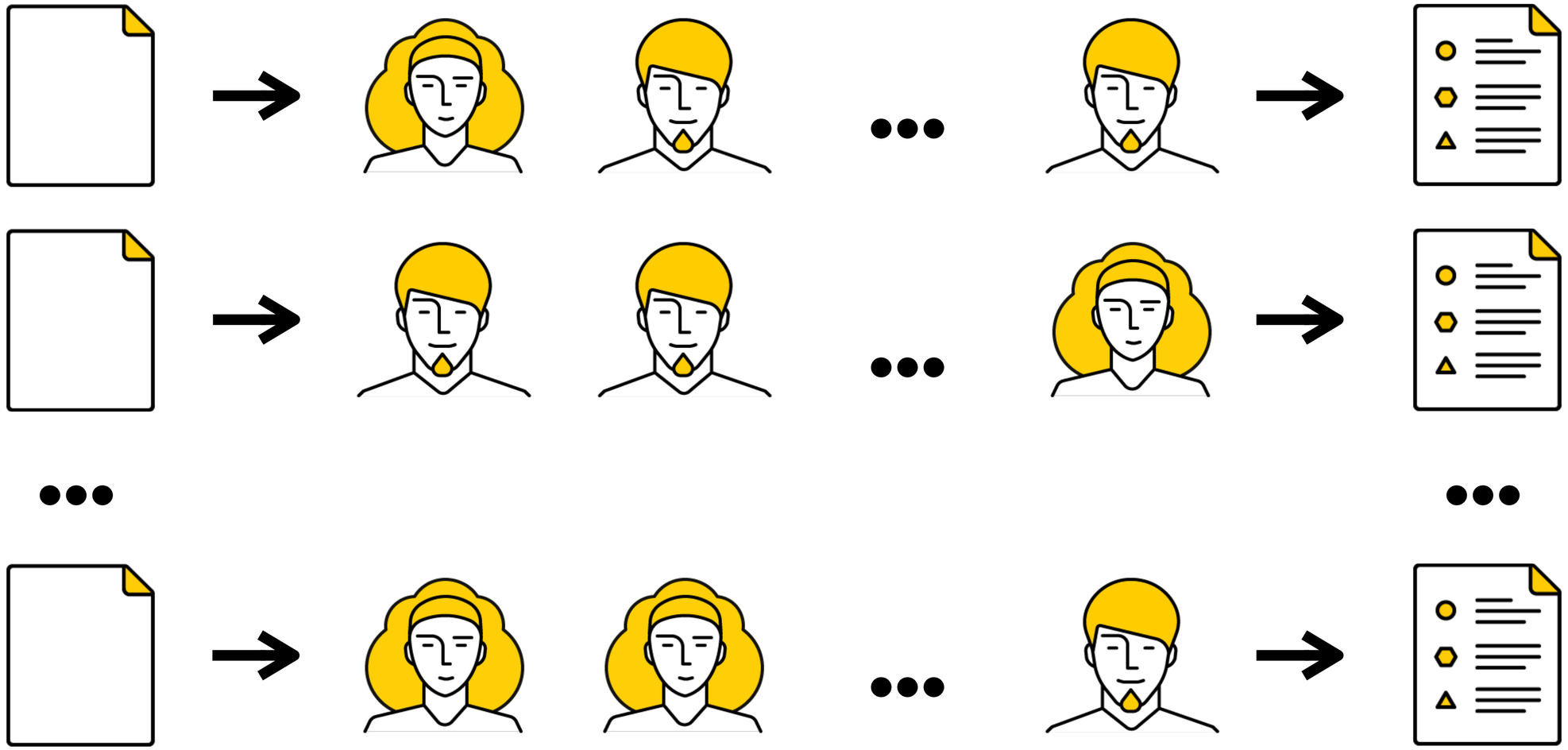
P, $\mathbf{d_j}$ L_{wj}, e_w



P, d_j **L_{wj}**, e_w



P, d_j L_{wj}, $\mathbf{e_w}$



Majority Vote

$F(L_{wj})$

- Учитываются только ответы исполнителей на задание в перекрытии
- «Побеждает» ответ, набравший большинство

$F(L_{wj}, e_w)$

- Взвешенный MV: ответ каждого исполнителя входит в итоговый с некоторым весом, пропорциональным качеству

Dawid-Skene

$F(L_{wj}, P, e_w)$

- Используется априорное распределение классов ответов
- Оценивается добросовестность исполнителей
- EM-алгоритм

Dawid and Skene, 1979 "Maximum Likelihood Estimation of Observer Error-rates using the EM Algorithm"

GLAD

Generative model of Labels, Abilities, and Difficulties

$F(L_{wj}, d_j, e_w)$

- Оценивается добросовестность исполнителей
- Дополнительно оценивается сложность каждого задания
- EM-алгоритм

Whitehill et al., 2009, "Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise"

Модель Андрея Мищенко

F (L_{wj}, d_j, p_w, q_w)

› Модификация модели GLAD

Модель Андрея Мищенко

$F(L_{wj}, d_j, p_w, q_w)$

- Модификация модели GLAD
- У исполнителей две характеристики:

Модель Андрея Мищенко

F (L_{wj}, d_j, p_w, q_w)

- Модификация модели GLAD
- У исполнителей две характеристики:

  Вероятность правильно ответить, если посмотрел на задание

Модель Андрея Мищенко

$F(L_{wj}, d_j, p_w, q_w)$

- Модификация модели GLAD
- У исполнителей две характеристики:

  Вероятность правильно ответить, если посмотрел на задание

  Вероятность, что посмотрел на задание

Наши наблюдения

Наши наблюдения

- Нет универсальной модели агрегации, которая давала бы статистически значимый прирост в качестве по сравнению с простым Majority Vote

Наши наблюдения

- Нет универсальной модели агрегации, которая давала бы статистически значимый прирост в качестве по сравнению с простым Majority Vote
- **Однако,** для каждого проекта в отдельности можно выбрать модель, использование которой даст до +15%% к качеству

Стоимость

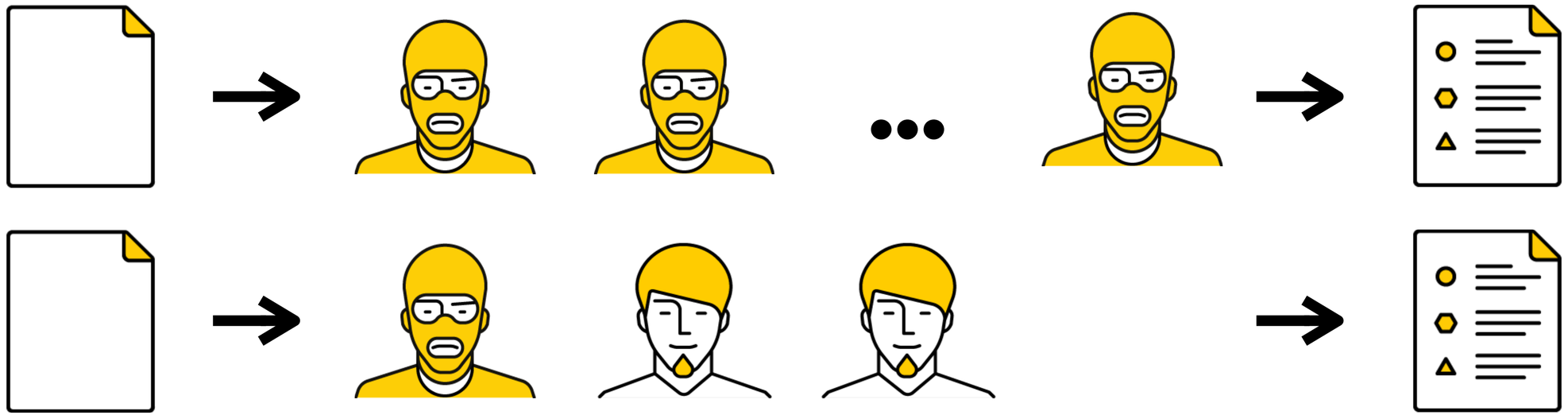
Динамическое перекрытие

Уменьшаем затраты без потерь в качестве

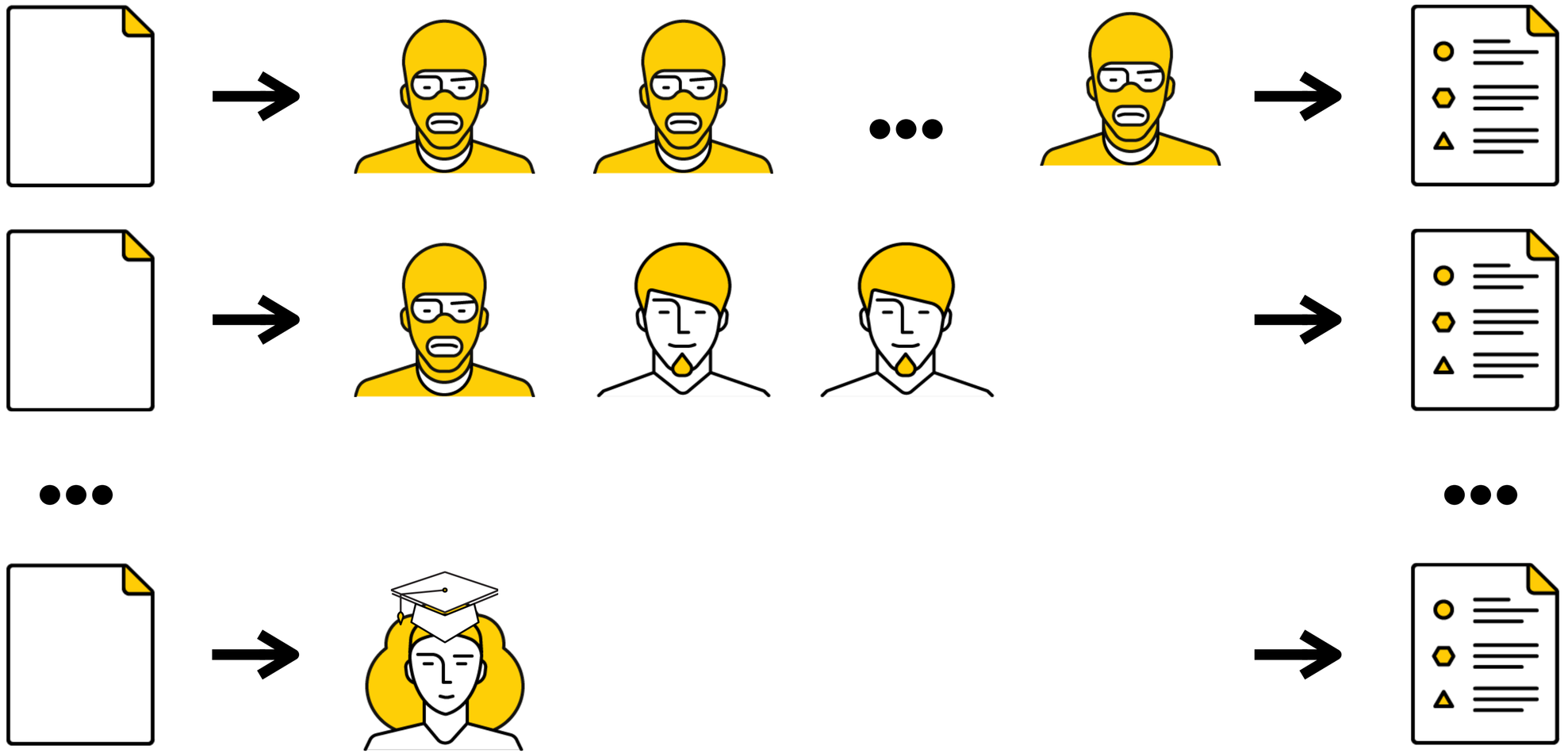
Динамическое перекрытие

Динамическое перекрытие

Динамическое перекрытие



Динамическое перекрытие



Экономия

- Динамическое перекрытие позволяет экономить до 40% бюджета без снижения качества

И снова качество

Куда вложить сэкономленное?

Куда вложить сэкономленное?

- Увеличение стоимости заданий не влечёт увеличения качества получаемых ответов

Куда вложить сэкономленное?

- Увеличение стоимости заданий не влечёт увеличения качества получаемых ответов

- Всё даже хуже...
  Высокая стоимость привлекает больше читеров

Теория игр

Теория игр

- Заказчик стремится максимизировать качество

Теория игр

- Заказчик стремится максимизировать качество
- Исполнитель стремится максимизировать заработок

Теория игр

- Заказчик стремится максимизировать качество
- Исполнитель стремится максимизировать заработок
- Равновесие достигается, когда плата за задание пропорциональна пользе, которую оно приносит для заказчика

Динамическое ценообразование

Мотивируем хороших исполнителей остаться
Увеличивается качество собираемых данных

Динамическое ценообразование

Динамическое ценообразование

- Цена за задание пропорциональна личному параметру точности исполнителя:
  — навык до 50: 0,01$
  — от 50 до 75: 0,02$
  — свыше 75: 0,05$

Динамическое ценообразование

- Цена за задание пропорциональна личному параметру точности исполнителя:
  — навык до 50: 0,01$
  — от 50 до 75: 0,02$
  — свыше 75: 0,05$
- Значимо повышает «время жизни» исполнителя на задании

В сухом остатке

В сухом остатке

- Краудсорсинг — это мощный инструмент для сбора данных

В сухом остатке

- Краудсорсинг — это мощный инструмент для сбора данных
- Практически любое задание (даже довольно сложное) можно построить на базе крауда

В сухом остатке

- Краудсорсинг — это мощный инструмент для сбора данных
- Практически любое задание (даже довольно сложное) можно построить на базе крауда
- Умные математические модели помогают улучшить качество собираемых данных и эффективно управлять бюджетом

Попробуйте Толоку для ваших задач!

https://toloka.yandex.com/for-requesters

Спасибо!

Артём Григорьев

Руководитель службы разработки

ortemij@yandex-team.ru