Yandex
Practice of Efficient Data Collection via Crowdsourcing: Aggregation, Incremental Relabelling, and Pricing

Alexey Drutsa, Valentina Fedorova, Dmitry Ustalov, Olga Megorskaya, Evfrosiniya Zerminova, Daria Baidakova
Part IV

Setting up and running label collection projects cont.

Daria Baidakova,
Project Manager

Yandex.Toloka is a service of Swiss company Yandex Services AG
Tutorial outline

Part I: 40 min
Main Components

Coffee break: 30 min

Part II: 25 min
Brainstorming pipeline

Introduction: 20 min

Part III: 10 min
Introduction to Crowd Platform

Lunch break: 90 min

Part IV: 85 min
Set & Run Projects

Part V: 35 min
Interface & Quality control

Part VI: 60 min
Set & Run Projects cont.

Part VI: 25 min
Theory on Aggregation

Coffee break: 30 min

Part VII: 20 min
Incremental relabeling and pricing

Part VIII: 10 min
Results & Conclusions
Practice starts: you have 60 min left

- You create 4 crowd projects
- You connect them in a pipeline
- You run them on real performers
- You have:
  › The instruction with all details
  › The dataset with photos

Tutorial chat with materials:
Most of us are at this step

**Project #1**
- Does a photo contain a specific item?
  - Create & Configure
  - Run
  - If yes
    - photos with items
  - If no
    - photos with items

**Tasks for performers:**
- a set of photos
- photos with items

**Project #2**
- Find a similar item in the online store
  - Create & Configure
  - Run

**Project #3**
- Is the found item (project 2) similar to the initial one? (post-verification)
  - Create & Configure
  - Run
  - If no
    - photos with incorrect similar items
  - If yes
    - photos with correct similar items

**Project #4**
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run

- Reject the result in Project #2
  - Do not pay for it
- Accept the result in Project #2
  - Pay for it

OR
- Run
Most of us are at this step

Project #1
- Does a photo contain a specific item?
  - Create & Configure
  - Run
  - If yes

Tasks for performers:
- a set of photos
- photos with items

Project #2
- Find a similar item in the online store
  - Create & Configure
  - Run

Project #3
- Is the found item (project 2) similar to the initial one? (post-verification)
  - Create & Configure
  - Run
  - If no
    - photos with incorrect similar items
  - If yes
    - photos with correct similar items

Project #4
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run

Options:
- OR
- Reject the result in Project #2
- Do not pay for it
- Accept the result in Project #2
- Pay for it
- photos w/ best substitute
Most of us are at this step

Project #1
- Does a photo contain a specific item?
  - Create & Configure
  - Run
- photos with items

Project #2
- Find a similar item in the online store
  - Create & Configure
  - Run
- photos with correct similar items

Project #3
- Is the found item (project 2) similar to the initial one? (post-verification)
  - Create & Configure
  - Run
- photos with incorrect similar items
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run
- photos w/ best substitute

Tasks for performers:
- a set of photos
- photos with items

If yes
- photos with best substitute

If no
- If yes
- Accept the result in Project #2
- Pay for it
- Run

If no
- Reject the result in Project #2
- Do not pay for it
- Run

OR
- Run
Most of us are at this step

Project #1
- Does a photo contain a specific item?
  - Create & Configure
  - Run
  - If yes

Project #2
- Find a similar item in the online store
  - Create & Configure
  - Run
  - If yes

Project #3
- Is the found item (project 2) similar to the initial one? (post-verification)
  - Create & Configure
  - Run
  - If no
  - If yes

Project #4
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run
  - OR

Tasks for performers:
- a set of photos
- photos with items
- photos with correct similar items
- photos with incorrect similar items
- photos w/ best substitute

Actions:
- If yes: Create & Configure
- If no: Reject the result in Project #2
  - Do not pay for it
  - OR
  - Accept the result in Project #2
    - Pay for it
Most of us are at this step

Project #1
- Does a photo contain a specific item?
  - Run
  - If yes
    - Create & Configure

Tasks for performers:
- a set of photos
- photos with items

Project #2
- Find a similar item in the online store
  - Run
  - Create & Configure

Project #3
- Is the found item (project 2) similar to the initial one? (post-verification)
  - Run
  - Create & Configure

Project #4
- Which of the found items is more similar to the initial one?
  - Run
  - Create & Configure

Create & Configure

- OR
- Reject the result in Project #2
- Do not pay for it
- Accept the result in Project #2
- Pay for it
Most of us are at this step

**Project #1**
- Does a photo contain a specific item?
  - Create & Configure
  - Run
  - If yes

**Project #2**
- Find a similar item in the online store
  - Create & Configure
  - Run
  - If yes

**Project #3**
- Is the found item (project 2) similar to the initial one? (post-verification)
  - Create & Configure
  - Run
  - If no
  - If yes

**Project #4**
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run
  - OR

Tasks for performers:
- A set of photos
- Photos with items
- Photos with incorrect similar items
- Photos with correct similar items
- Photos w/ best substitute

- Reject the result in Project #2
  - Do not pay for it
- Accept the result in Project #2
  - Pay for it
Most of us are at this step

Project #1
- Does a photo contain a specific item?
- Create & Configure
- Run
- If yes: photos with items
- If no: a set of photos

Tasks for performers:
- a set of photos

Project #2
- Find a similar item in the online store
- Create & Configure
- Run
- If yes: photos with similar items
- If no: photos with incorrect similar items

Project #3
- Is the found item (project 2) similar to the initial one? (post-verification)
- Create & Configure
- Run
- If yes: photos with correct similar items
- If no: photos with best substitute

Project #4
- Which of the found items is more similar to the initial one?
- Create & Configure
- Run
- OR

Create & Configure
- Run
- If yes: Pay for it
- If no: Do not pay for it

Reject the result in Project #2
- Do not pay for it

Accept the result in Project #2
- Pay for it
Most of us are at this step

**Project #1**
- Does a photo contain a specific item?
- **Create & Configure**
- **Run**
- If yes

**Project #2**
- Find a similar item in the online store
- **Create & Configure**
- **Run**
- **Create & Configure**
- **Run**
- **Create & Configure**

**Project #3**
- Is the found item (project 2) similar to the initial one? (post-verification)
- **Create & Configure**
- **Run**
- If no
- If yes

**Project #4**
- Which of the found items is more similar to the initial one?
- **Create & Configure**
- **Run**

**Tasks for performers:**
- a set of photos
- photos with items

**Create & Configure**
- **Run**
- If yes

**Create & Configure**
- **Run**
- **Create & Configure**

**Create & Configure**
- **Run**
- **Create & Configure**

**Create & Configure**
- **Run**
- **Create & Configure**

**OR**
- photos w/ best substitute

**Reject the result in Project #2**
- Do not pay for it

**Accept the result in Project #2**
- Pay for it
Thank you!
Questions?

Daria Baidakova
Project Manager

dbaidakova@yandex-team.ru

https://research.yandex.com/tutorials/crowd/kdd-2019