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

Daria Baidakova,
Project Manager

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

Part I: 40 min
Main Components

Part II: 25 min
Brainstorming pipeline

Coffee break: 30 min

Part III: 10 min
Introduction to Crowd Platform

Part IV: 85 min
Set & Run Projects

Lunch break: 90 min

Part V: 35 min
Interface & Quality control

Part VI: 60 min
Set & Run Projects cont.

Coffee break: 30 min

Part VII: 20 min
Incremental relabeling and pricing

Part VIII: 10 min
Results & Conclusions
What you need for the practical session

We are starting the practical session

We give you a card with information and links to:
› A step-by-step instruction to configure and run our crowd projects
› A dataset with photos that should be labeled
› Login+Password to sign in Yandex.Toloka as a requester

We also provide several copies of a printed version of the instruction

Did everybody receive this card?
Requester account that you received

You have Login+Password to sign in Yandex.Toloka as a requester

The same account is given for several participants (a group)

› So, you can divide work on the project configuration within this group
› Or, each member of a group may work individually and create the whole pipeline by her/himself
Sign in Yandex.Toloka as a requester

1. Go to [https://ya.cc/tlk](https://ya.cc/tlk)
2. Click on “Sign in” in the top-right corner
3. Use received Login+Password to sign in
Requester account that you received

You have Login+Password to sign in Yandex.Toloka as a requester

- The account of this requester has money
  ▶ So, you will run your tasks on real crowd performers!
Practice: creating a real crowdsourcing pipeline

Now we will create a real simplified crowdsourcing pipeline

To simplify the task, we ask you to:
› Finding a substitute for **one type** of item
› Choose any item you want to find the best substitute for. For example, **Shoes**
Reminder: we implement and run our pipeline

Tasks for performers:

- a set of photos
- photos with items
- photos with similar items
- photos with incorrect similar items
- photos w/ best substitute

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

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

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

OR
- photos w/ best substitute
You can divide work within a participant group

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

**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

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

### Can do in parallel

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

**Create & Configure**
- photos with correct similar items
- photos with incorrect similar items

**Run**
- reject the result in Project #2
  - Do not pay for it
- Pay for it

**OR**
- photos w/ best substitute
Step #1

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

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

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

Project #3
- Is the found item (project 2) similar to the initial one? (post-verification)
- If yes: Create & Configure
- Run
- photos w/ best substitute
- Which of the found items is more similar to the initial one?
- If yes: Accept the result in Project #2
- Pay for it
- If no: Reject the result in Project #2
- Do not pay for it

Project #4
- Create & Configure
- Run

OR
- photos w/ best substitute
Step #2

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

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

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

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

- Photos w/ best substitute
- Photos with correct similar items
- Photos with incorrect similar items

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

- Accept the result in Project #2
  - Pay for it
Step #3

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

**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
  - photos w/ best substitute

- **Reject the result in Project #2**
  - Do not pay for it
- **Accept the result in Project #2**
  - Pay for it
Step #4

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:

Project #2

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

If yes
- photos with correct 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 no
- Run

If yes
- photos w/ best substitute

Project #4

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

OR
- Reject the result in Project #2
  - Do not pay for it
- Accept the result in Project #2
  - Pay for it
Step #5

**Project #1**
- Does a photo contain a specific item?
  - Yes: Create & Configure
  - No: Run

**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)
  - Yes: Create & Configure
  - No: Run

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

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

- If yes:
- If no:
- If yes:
- OR

- Accept the result in Project #2
- Pay for it
- Reject the result in Project #2
- Do not pay for it
Step #6

**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
  - Create & Configure
  - Run
  - If no

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

**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

If yes:
- photos with correct similar items
  - Accept the result in Project #2
  - Pay for it

If no:
- photos with incorrect similar items
  - Reject the result in Project #2
  - Do not pay for it

OR
- photos w/ best substitute
Step #7

**Project #1**
- Does a photo contain a specific item?
  - Yes: Create & Configure
  - No: Run

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

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

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

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

**Create & Configure**
- Run

**Options**
- Reject the result in Project #2
- Pay for it

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

**OR**
- photos w/ best substitute
Step #8

Tasks for performers:

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

 Photos w/ best substitute

Run

Photos with correct similar items

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

Photos with incorrect similar items

Accept the result in Project #2
Pay for it
Practice pt 1 starts

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

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 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
- photos with similar items
- photos with incorrect similar items
- photos with best substitute

- Run
- Create & Configure
- 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?
  - Yes: Create & Configure
  - No: Run

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

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

Flow:
- If yes: photos with best substitute
- If no: Reject the result in Project #2
- Accept the result in Project #2
- Pay for it
- Do not pay for it
Most of us are at this step

Project #1

- Does a photo contain a specific item?
- Create & Configure
- Run

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:
- If yes:

Project #4

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

OR

- Photos with incorrect similar items
- Reject the result in Project #2
- Do not pay for it

- Photos with correct similar items
- 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**
  - If yes: **Create & Configure**

**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: **Create & Configure**
  - If yes: **Create & Configure**

**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

- **Create & Configure**
- **Run**
- If yes: **Create & Configure**
- If yes: **Create & Configure**
- If no: **Create & Configure**
- If yes: **Create & Configure**

- **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

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

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

- photos with similar items
- photos with incorrect similar items

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

- photos with correct similar items
- Reject the result in Project #2
  - Do not pay for it
- Accept the result in Project #2
  - Pay for it

Project #4
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run
  - OR
    - photos w/ best substitute
Most of us are at this step

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

**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)
  - If no: Run
  - If yes: Create & Configure

**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
- photos with correct similar items
- photos with incorrect similar items
- photos w/ best substitute

If yes:
- Create & Configure
- Run

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

If yes:
- Accept the result in Project #2
- Pay for it

OR
- Accept the result in Project #2
- Pay for it

---

Does a photo contain a specific item?

Find a similar item in the online store

Is the found item (project 2) similar to the initial one? (post-verification)

Which of the found items is more similar to the initial one?
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
  - If yes

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

- photos with incorrect similar items
- photos with correct similar items

Project #4
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run
  - 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
Most of us are at this step

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

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

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

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

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

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
Thank you!
Questions?

Daria Baidakova
Project Manager

dbaidakova@yandex-team.ru
https://research.yandex.com/tutorials/crowd/kdd-2019