Yandex
Crowdsourcing Practice for Efficient Data Labeling: Aggregation, Incremental Relabeling, and Pricing

Alexey Drutsa, Valentina Fedorova, Dmitry Ustalov, Olga Megorskaya, Evfrosiniya Zerminova, Daria Baidakova
Setting up and running label collection projects

Daria Baidakova, Project Manager

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

**Introduction:** 15 min

**Part I:** 20 min  
Main Components

**Part II:** 10 min  
Introduction to Crowd Platform

**Part III:** 15 min  
Brainstorming pipeline

**Part IV:** 60 min  
Set & Run Projects

**Part V:** 25 min  
Theory on Aggregation

**Break:** 30 min

**Part VI:** 20 min  
Set & Run Projects cont.

**Part VII:** 10 min  
Results & Conclusions
What you need for the practice session

We are starting the practice session

We give you 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
Sign in Yandex.Toloka as a requester

1. Go to https://toloka.ai
2. Click on “Sign in” in the top-right corner
3. Use received Login+Password to sign in
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

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
  - If yes

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
Step #1

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

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

---

If yes
- Create & Configure
- Run

If no
- Create & Configure
- Run

If yes
- Create & Configure
- Run

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

OR
- photos w/ best substitute
Step #2

**Tasks for performers:**
- a set of photos
- photos with items
- photos with similar items
- photos with incorrect similar items
- photos with the best substitute

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

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

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

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

- **OR**
  - **Accept the result in Project #2**
  - **Pay for it**
- **Reject the result in Project #2**
- **Do not pay for it**
Step #3

**Project #1**

Does a photo contain a specific item?
- Create & Configure
- Run
  - If yes
    - photos with items

**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 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**

Which of the found items is more similar to the initial one?
- Create & Configure
- Run
  - OR
    - photos w/ best substitute
Step #4

**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?
- **Create & Configure**
- **Run**
- If no

**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

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

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

- photos w/ best substitute
Step #5

Tasks for performers:

- A set of photos
- Photos with items

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

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

Options:
- Reject the result in Project #2
- Do not pay for it
- Accept the result in Project #2
- Pay for it
- OR
- Photos w/ best substitute
Step #6

**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

**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
    - Reject the result in Project #2
    - Accept the result in Project #2
      - Pay for it
      - Do not 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
Step #7

Project #1

Does a photo contain a specific item?

- Create & Configure
- Run

If yes

Photos with items

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

Does a photo contain a specific item?

- Create & Configure
- Run

If no

Photos with incorrect similar items

If yes

Photos with correct similar items

Project #3

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

- Create & Configure
- Run

If yes

Photos with best substitute

Accept the result in Project #2
Pay for it

If no

Reject the result in Project #2
Do not 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
**Step #8**

**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

**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
        - 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

**Project #4**
- Which of the found items is more similar to the initial one?
  - Create & Configure
  - Run
  - OR
  - Photos w/ best substitute
In this part we create 2 projects:

- You configure crowd projects
- You connect them in a pipeline
- You run them on real performers

You have:
- The instruction with all details
  https://tlk.s3.yandex.net/relsubstitutes/SIGMOD_instruction.pdf
- The dataset with photos
  http://tlk.s3.yandex.net/relsubstitutes/dataset_Y.tsv
  where Y is any number from 1 to 10
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

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
- photos with correct similar items
- photos with incorrect 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

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

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

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

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

If no
- Create & Configure
- 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

**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

- If yes
- photos with correct similar items
- photos with incorrect similar items
- photos w/ best substitute

- 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

Run
Run
Run
Run
Most of us are at this step

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

**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:
  - Create & Configure
  - Run

- OR

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

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

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
  - 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
  - Run

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

OR

- photos w/ best substitute
Thank you!
Questions?

Daria Baidakova
Project Manager

dbaidakova@yandex-team.ru

https://research.yandex.com/tutorials/crowd/sigmod-2020