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

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

Introduction to Yandex.Toloka for requesters

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

**Break: 30 min**

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

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

**Part VII: 10 min**
Results & Conclusions
Requester's dashboard

- Requester name to show performers
- Pending amount for tasks waiting for acceptance
- Your balance
- Help and useful links
- Your projects will be listed here

Yandex Toloka

Projects Users Skills Profile Messages

Projects (Active)

- No data

- Find project

- Create project

- Your projects will be listed here
How Toloka works

Source data

- Images
- Texts
- Audio
- Videos
- Addresses
- Links

Task

- Project
- Pool
- TSV file

Performers

Results
Tasks in Toloka

**Project**

- Instructions
  - examples
  - acceptance criteria
  - steps to follow

- Specification
  - input
  - output

- Task interface
  - HTML
  - JavaScript
  - CSS

**Pool**

- Parameters
  - price
  - overlap
  - acceptance type
  - time to complete

- User filters
  - language and region
  - gender and age
  - device and OS
  - skill

- Quality control
  - fast responses
  - captcha
  - majority vote
  - control tasks

**TSV file**

- Data to use in the task
  - texts
  - links to audio, video or image files
  - correct responses
  - hints for training
## Task templates

Choose one or create your own

<table>
<thead>
<tr>
<th>Task Template</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offline data collection</td>
<td>Ask performers to complete tasks in the mobile app: collect data in the field, take photos, or record audio</td>
</tr>
<tr>
<td>Data search</td>
<td>Collect and annotate any type of information found online</td>
</tr>
<tr>
<td>Side-by-side comparison</td>
<td>Users vote for the best option and explain why</td>
</tr>
<tr>
<td>Image or voice recognition</td>
<td>Get human feedback when automatic recognition fails</td>
</tr>
<tr>
<td>Data classification</td>
<td>Categorize large amounts of data</td>
</tr>
<tr>
<td>Content moderation</td>
<td>Check if content meets your guidelines</td>
</tr>
<tr>
<td>Content generation</td>
<td>Users submit a photo, audio recording, or original text</td>
</tr>
<tr>
<td>Relevance assessment</td>
<td>Evaluate results for search queries</td>
</tr>
<tr>
<td>Survey participation</td>
<td>Get responses from a large audience</td>
</tr>
</tbody>
</table>
Creating a project

**Input data** is what we show the performer in the task: links to images, text, video.

**Output data** is what we get back from the performer as a result of task completion: text, url, photo.

You can use **JavaScript** to configure validation of responses or show follow-up questions.

**Task interface** defines what the task looks like for the performer.
Pool settings

Parameters
- Price
- Time to complete
- Overlap
- Acceptance type

User filters
- Language and region
- Age
- Gender
- Device and OS
- Skill

Quality control
- Fast responses
- Captcha
- Majority vote
- Control tasks
- Skipped assignments
and more…
TSV file with tasks

Create a TSV file with links to images, audio, video, texts, or coordinates

Upload the TSV file to the pool

Start the pool to assign tasks to performers
**Task types**

- **Main**: Tasks that you don't have answers for, and you want to get responses from performers.
  - INPUT:image

- **Control (golden)**: Tasks that you know the correct answers to. Used for checking performers.
  - INPUT:image
  - GOLDEN:result

- **Training**: Tasks with correct answers and hints. Used for training performers to complete the task.
  - INPUT:image
  - GOLDEN:result
  - HINT:text
Task completion

Track progress on your tasks on the project page, in the pool, and in Statistics.
Results

TSV file with responses

Aggregated results

Photo, video, audio
Thank you!

Questions?

Daria Baidakova
Project Manager

dbaidakova@yandex-team.ru

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