

**Y**andex

# **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 III

# Introduction to Yandex.Toloka for requesters

Evfrosiniya Zerminova,  
Head of Data Analysis and Research Group

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

# Tutorial outline

**Introduction: 20 min**

**Part I: 40 min**  
Main Components

**Coffee break:**  
30 min

**Part II: 25 min**  
Brainstorming  
pipeline

**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: 25 min**  
Theory on  
Aggregation

**Coffee break:**  
30 min

**Part VI: 60 min**  
Set & Run Projects  
cont.

**Part VII: 20 min**  
Incremental  
relabeling and pricing

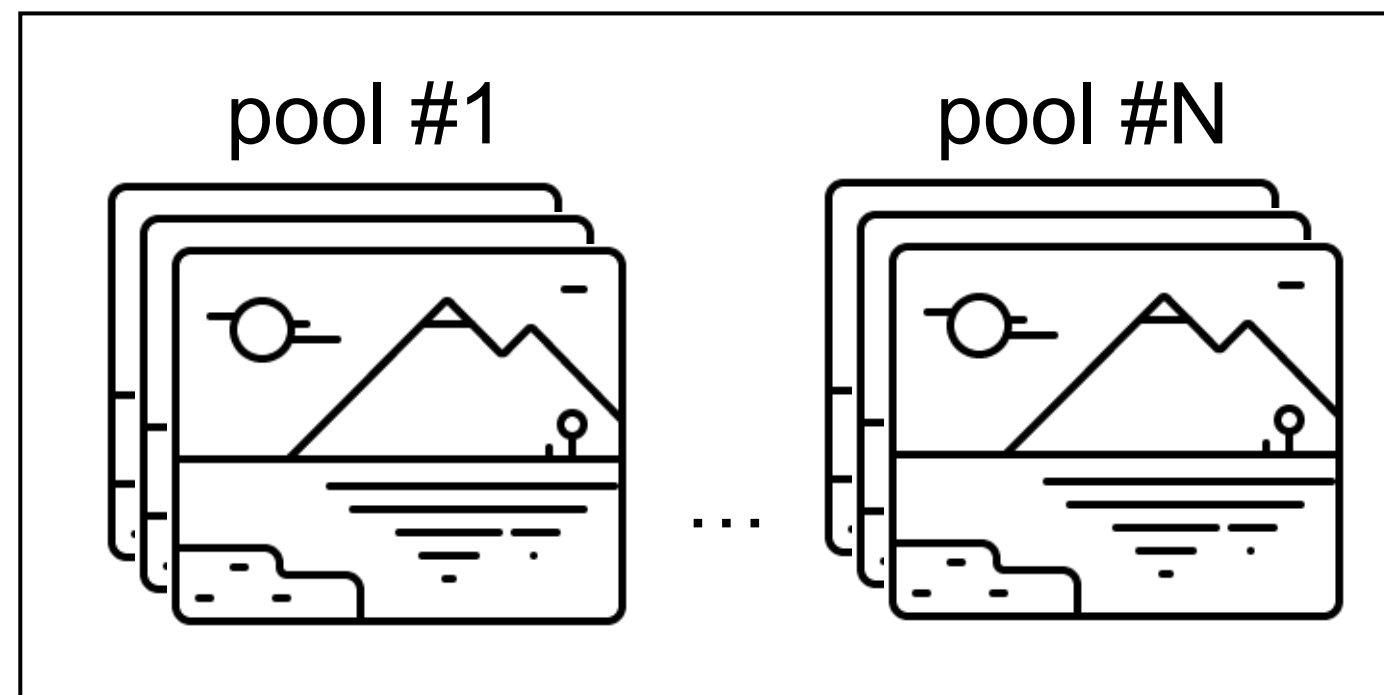
**Part VIII: 10 min**  
Results &  
Conclusions



# Key types of instances in Yandex.Toloka

## Project

- › Defines the structure of tasks
- › Defines how to perform them

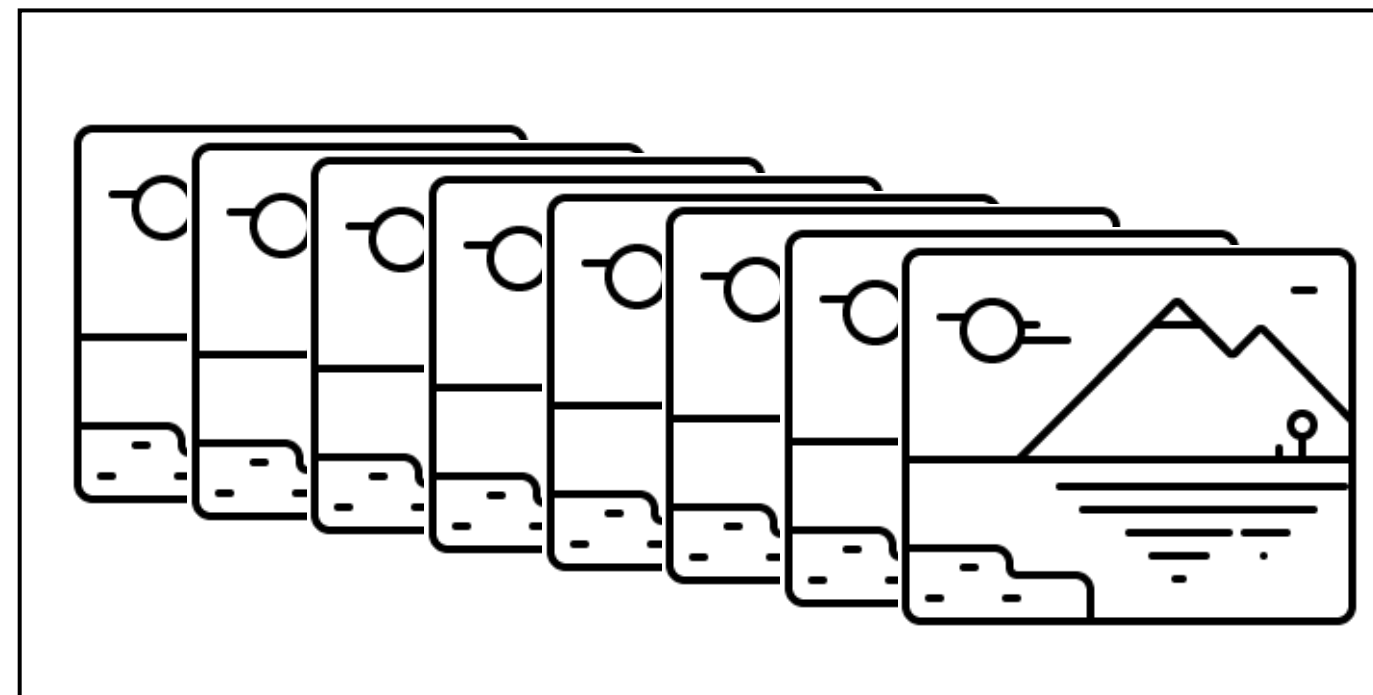


### Configure in a project:

- › Input and output data types
- › Task interface
- › Task instruction

## Pool

- › Is a batch of tasks
- › Defines access of performers

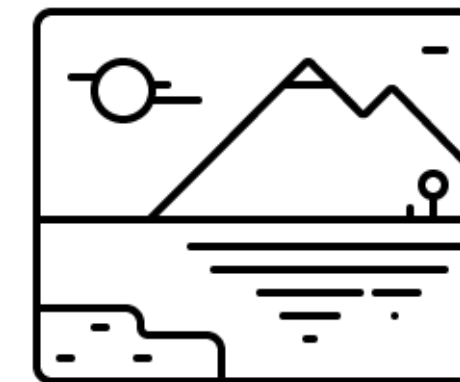


### Configure in a pool:

- › Performer filters
- › Quality control mechanisms
- › Overlap settings

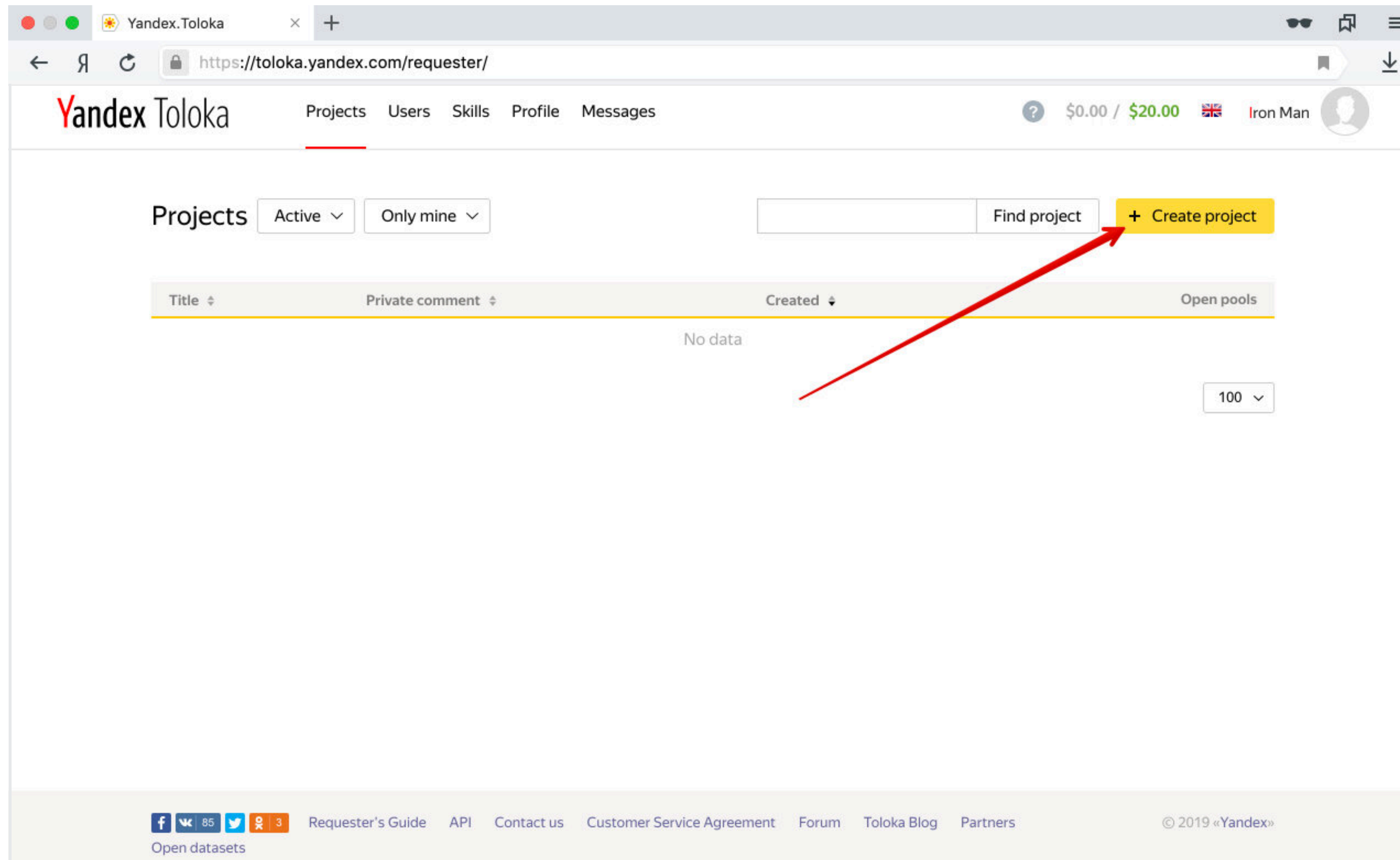
## Task

- › A particular input data
- › Results for it from performers



# **Project: creation & configuration**

# Project: creation



# Project: configure name and instruction

Yandex.Toloka

+

← ↻ 🔍

https://toloka.yandex.com/requester/new/project?templateId=blank

🔖 ⬇

PROJECT NAME

Are there shoes in the picture?

✕

DESCRIPTION

Look at the picture and tell whether there are shoes in it.

✕

INSTRUCTIONS ?

🔠 🎨 **B** *I* U 🖼️ 🔗 ☰ ☷ ☰ ☷ ☷ ☷

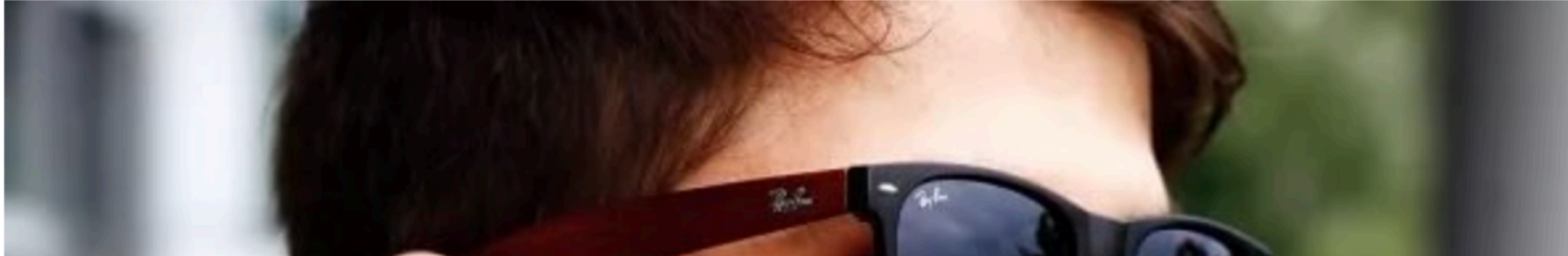
📐 👁 <>

Look at the photo and decide, whether there are **shoes** in the photo.

If there are shoes in the picture, click "YES"

If there are no shoes, click "NO"

For example, there are shoes in the photo. Therefore, the correct answer is "YES"



PRIVATE COMMENT

for private notes, not visible to users

# Project: configure in/out data types

Yandex.Toloka

https://toloka.yandex.com/requester/new/project?templateId=categorize

SPECIFICATIONS ?

Input data

Title: image

Required: yes

Type: URL

Add field

Output data

Title: result

Required: yes

Type: string

Title: like

Required: yes

Type: boolean

Add field

# Project: configure interface

The screenshot shows the Yandex.Toloka web interface for configuring a new project. The browser address bar shows the URL: `https://toloka.yandex.com/requester/new/project?templateId=categorize`. The main area is titled "TASK INTERFACE" and contains a "Preview" button and a code editor. A red arrow points to the "TASK INTERFACE" label. The code editor is divided into three sections: HTML, JavaScript, and CSS.

**HTML:**

```
1 {{img src=image width="100%" height="400px"}}
2
3 {{field type="radio" name="result" value="OK" label="Good" hotkey="1"}}
4 {{field type="radio" name="result" value="BAD" label="Bad" hotkey="2"}}
5 {{field type="radio" name="result" value="404" label="Failed to load"
6 hotkey="3"}}
7 <br>
8 {{field type="checkbox" name="like" label="Do you like the photo?"
9 hotkey="q"}}
```

**JavaScript:**

```
1 exports.Task = extend(TolokaHandlebarsTask, function (options) {
2   TolokaHandlebarsTask.call(this, options);
3 });
```

**CSS:**

```
1
```

At the bottom of the interface, there are three buttons: "Cancel", "Save", and "Save and continue".



# Project: use the task preview to see how it works

Yandex.Toloka

Yandex.Toloka

+

https://toloka.yandex.com/requester/project/preview

Project preview

Change input data

Tasks

In progress


Messages

What is the cat's mood?

?

\$0.00 / \$0.00

Instructions




1 ☐ Good

2 ☐ Bad

3 ☐ Failed to load

q ☐ Do you like the photo?




1 ☐ Good

2 ☐ Bad

3 ☐ Failed to load

q ☐ Do you like the photo?



1 ☐ Good

2 ☐ Bad

3 ☐ Failed to load

q ☐ Do you like the photo?

Exit

Skip

Submit

# Project: saving

The screenshot shows the Yandex.Toloka web interface for configuring a new project. The browser address bar displays `https://toloka.yandex.com/requester/new/project?templateId=categorize`. The main content area is titled 'TASK INTERFACE' and contains three tabs: 'Preview', 'HTML', 'JS', and 'CSS'. The 'HTML' tab is active, showing the following code:

```
1 {{img src=image width="100%" height="400px"}}
2
3 {{field type="radio" name="result" value="OK" label="Good" hotkey="1"}}
4 {{field type="radio" name="result" value="BAD" label="Bad" hotkey="2"}}
5 {{field type="radio" name="result" value="404" label="Failed to load"
6 hotkey="3"}}
7 <br>
8 {{field type="checkbox" name="like" label="Do you like the photo?"
9 hotkey="q"}}
```

The 'JS' tab is also visible, showing the following code:

```
1 exports.Task = extend(TolokaHandlebarsTask, function (options) {
2   TolokaHandlebarsTask.call(this, options);
3 });
```

The 'CSS' tab is empty. At the bottom of the interface, there are three buttons: 'Cancel', 'Save', and 'Save and continue'. A red arrow points from the 'HTML' tab area to the 'Save' button.



# **Pool: creation & configuration**

# Pool: creation

Yandex Toloka

ProjectsUsersSkillsProfileMessages

?

\$0.00 / \$20.00

🇬🇧

Iron Man

Are there shoes in the picture? — active

Project actions ▾

Statistics for 7 days

Submitted tasks	Spent	Quality: control tasks	Quality: training tasks	Average submit time	Users	Banned users
0	0\$	-	-	-	0	0

PoolsTrainingStatisticsQuality control

Search

Filters

Add a pool

Title ▾	Priority ▾	Progress	Status ▾	Started ▾	To be completed
Add a pool to upload tasks, filter users, and set quality control rules.					

50 ▾

f

vk

85

tw

3

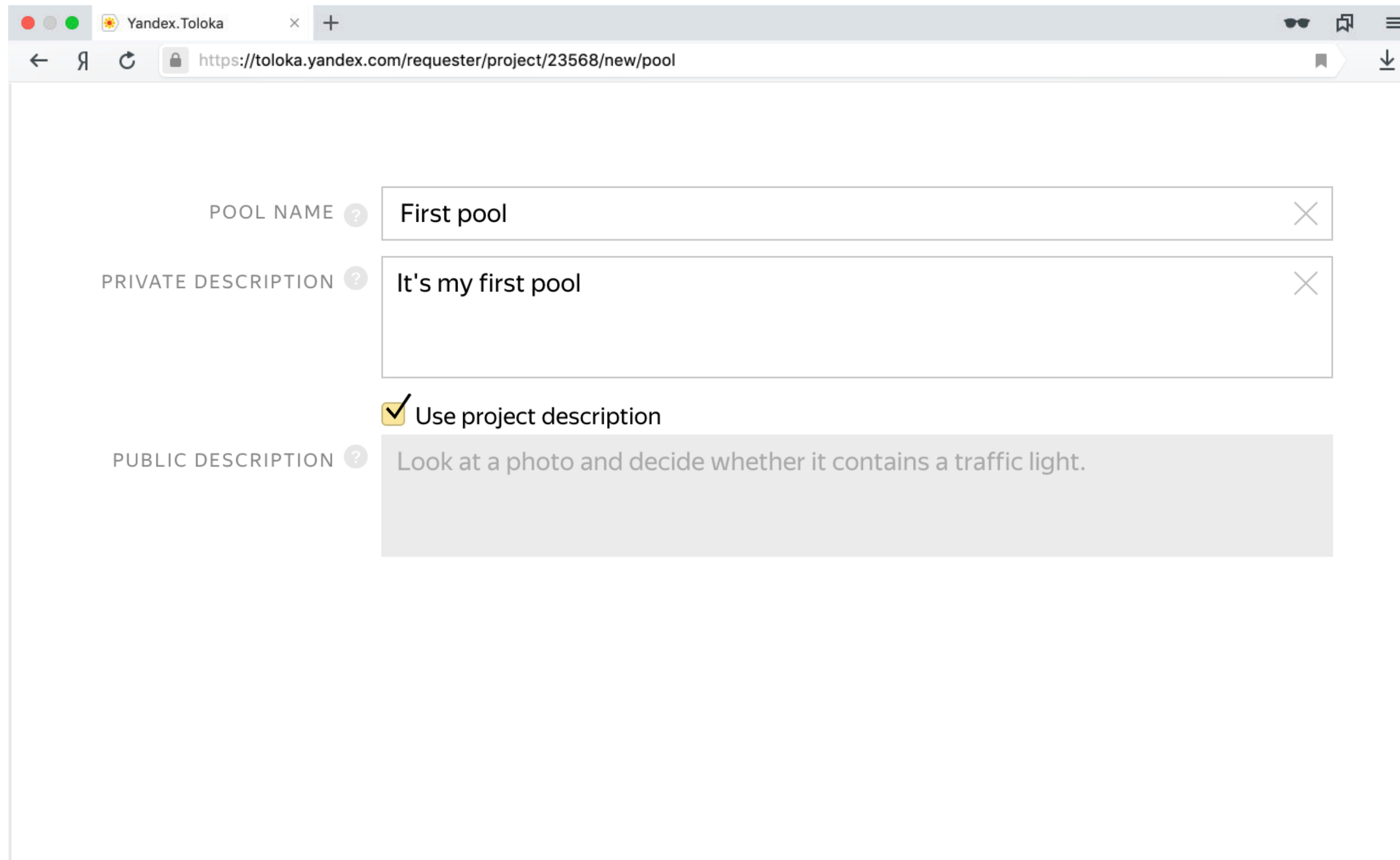
3

Requester's GuideAPIContact usCustomer Service AgreementForumToloka BlogPartners

Open datasets

© 2019 «Yandex»

# Pool: configure name and description



Yandex.Toloka

https://toloka.yandex.com/requester/project/23568/new/pool

POOL NAME ? First pool

PRIVATE DESCRIPTION ? It's my first pool

☒ Use project description

PUBLIC DESCRIPTION ? Look at a photo and decide whether it contains a traffic light.

# Pool: configure pricing and overlap

Yandex.Toloka

https://toloka.yandex.com/requester/project/23568/new/pool

Price per task suite

PRICE ?

0.01

×

MARKUP ?

0.005

+

Dynamic pricing

Overlap

OVERLAP ?

3

×

DYNAMIC OVERLAP ?

Off

# Pool: configure timing and post verification

Yandex.Toloka

https://toloka.yandex.com/requester/project/23568/new/pool

Task settings

TIME ON TASK ?

600

CAPTCHA FREQUENCY ?

None

KEEP TASK ORDER ?

No

ADULT CONTENT ?

Yes

EXPIRES ?

2021-01-29

TIME TO CLOSE ?

0

PRIORITY ?

0

NON-AUTOMATIC ACCEPTANCE ?

No

REVIEW PERIOD

# Pool: filter performers by their profiles

Yandex.Toloka

+

https://toloka.yandex.com/requester/project/23568/new/pool

PROFILE

Adult content

On

+

AND

PROFILE

Birth date

=

05.09.1990

+

AND

PROFILE

City

≠

Anchorage

+

AND

PROFILE

Education

=

Secondary

+

AND

PROFILE

Languages

=

English

+

# Pool: filter performers by computed properties

Yandex.Toloka

https://toloka.yandex.com/requester/project/23568/new/pool

COMPUTED

Browser = YANDEX\_BROWSER

AND

COMPUTED

Client = Toloka web version

# Pool: filter performers by custom skills

Yandex.Toloka

+

https://toloka.yandex.com/requester/project/23568/new/pool

User filter ?

Copy settings from...

Add filter

Create skill

SKILLS

Data search

>

10

+



# Pool: control quality via a rule on golden set

Yandex.Toloka × +

← ↻ 🔒 <https://toloka.yandex.com/requester/project/23568/new/pool> 📖 ⬇

CONTROL TASKS ?

History size

If

and

then

# Pool: control quality via a rule on majority vote

Yandex.Toloka × +

← ↻ 🔒 https://toloka.yandex.com/requester/project/23568/new/pool

MAJORITY VOTE ?

Accept as majority  ×

History size

If  ▾   × +

and  ▾   × 🗑

then  ▾  ▾  ×

×

+

# Pool: control quality via a rule on post verification

Yandex.Toloka × +

← ↻ 🔒 <https://toloka.yandex.com/requester/project/23568/new/pool> 📖 ⬇

RECOMPLETION OF REJECTED ASSIGNMENTS ?

History size  🗑

If  ▾   ✕ +

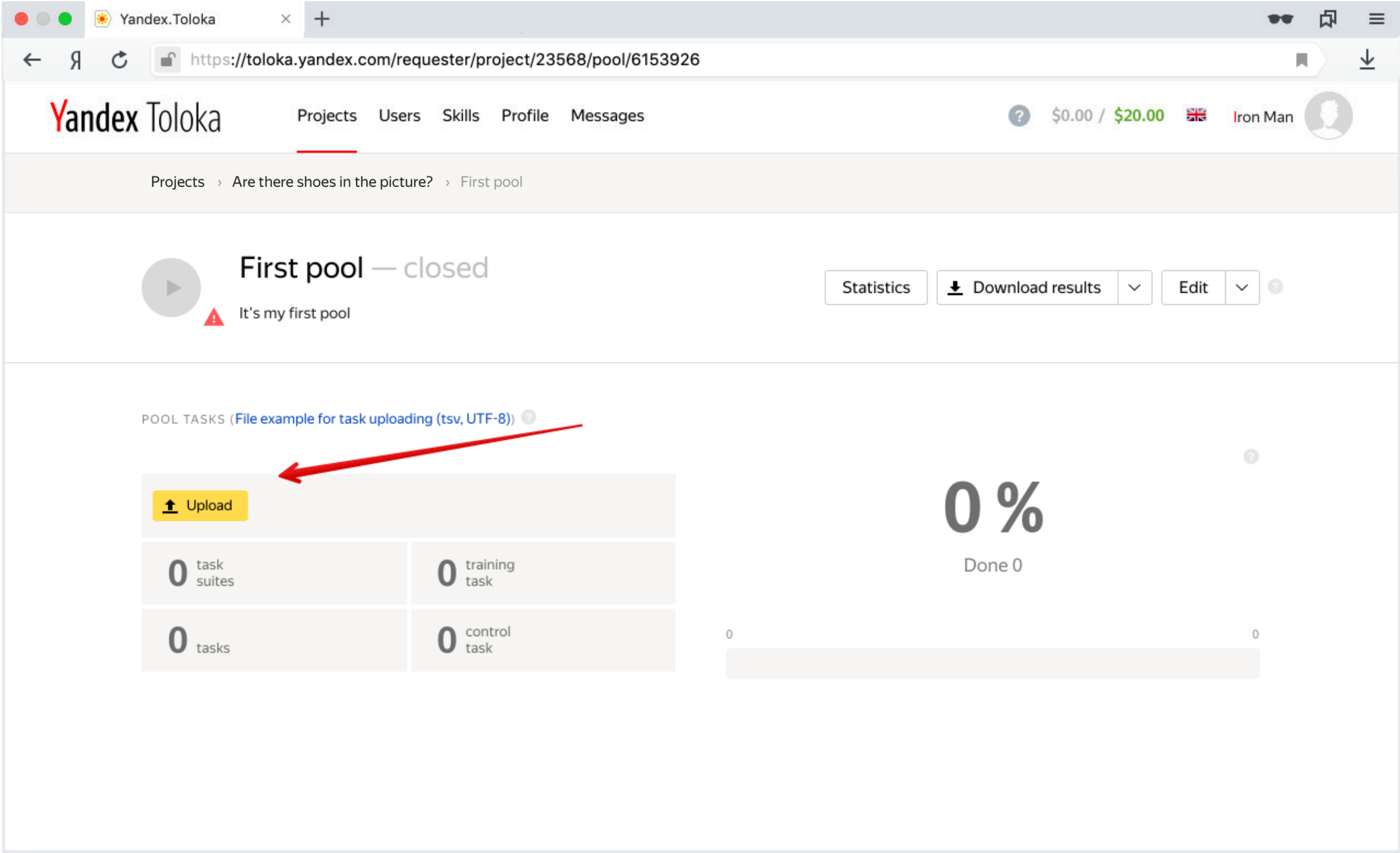
and  ▾   ✕ 🗑

then  ▾  ▾  ✕

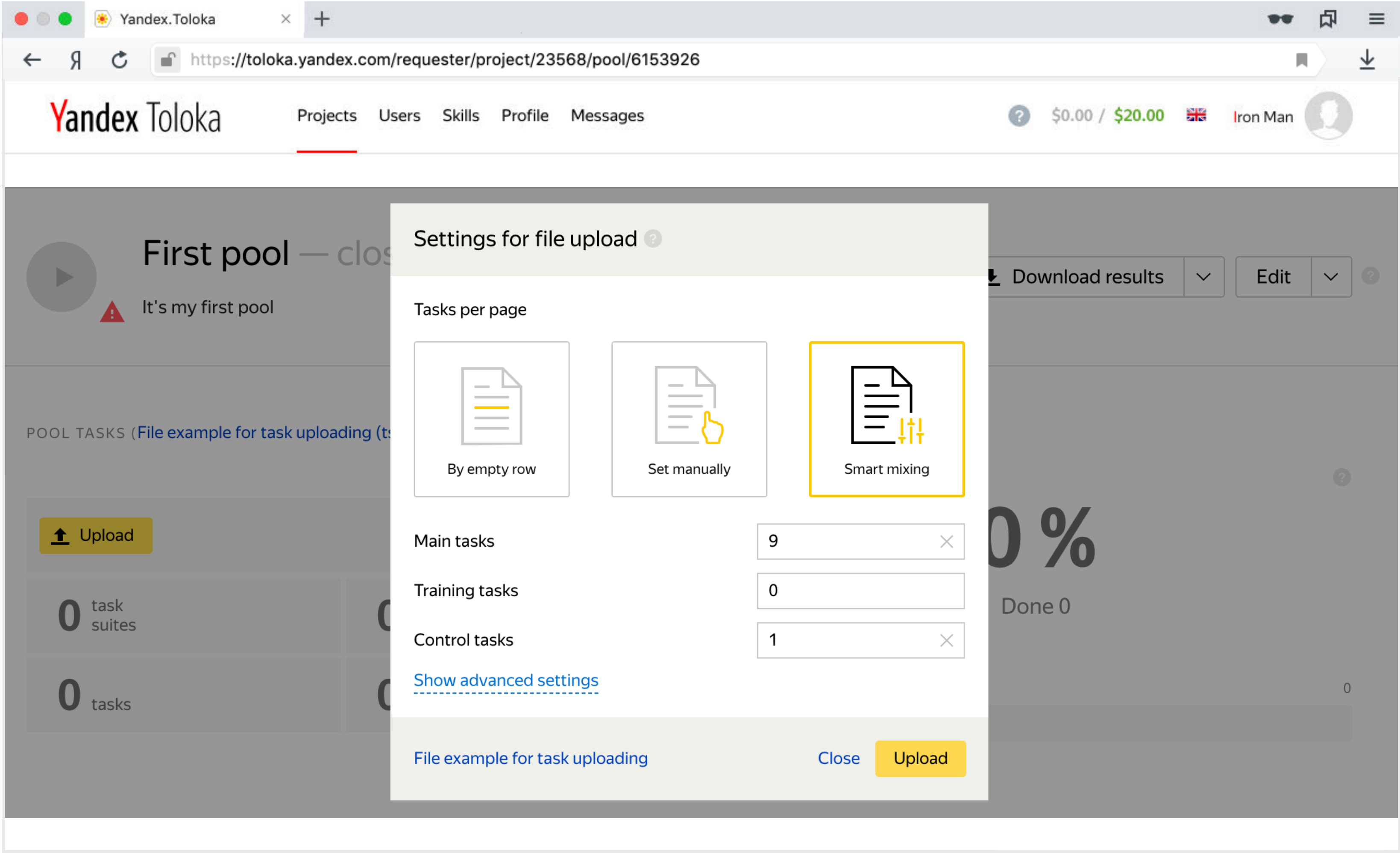
✕ +

# **Task uploading & golden set creation**

# Task: uploading



# Task: uploading



# Task: edit for creation of control tasks

The screenshot shows the Yandex Toloka web interface. The browser address bar displays the URL `https://toloka.yandex.com/requester/project/23568/pool/6153926`. The page header includes the Yandex Toloka logo, navigation links (Projects, Users, Skills, Profile, Messages), and user information (\$0.00 / \$20.00, Iron Man). The main content area shows a task pool titled "First pool — closed" with a subtitle "It's my first pool". Below this, there are buttons for "Statistics", "Download results", and "Edit". The "Edit" button is highlighted with a red arrow. The task pool statistics show 0 task suites, 0 training tasks, 100 tasks, and 0 control tasks. A progress bar indicates 0% completion.

Yandex Toloka Projects Users Skills Profile Messages \$0.00 / \$20.00 Iron Man

Projects > Are there shoes in the picture? > First pool

First pool — closed  
It's my first pool

Statistics Download results Edit

POOL TASKS (File example for task uploading (tsv, UTF-8))

Upload files Delete Edit

0 task suites 0 training task

100 tasks 0 control task

0 %  
Done 0

# Task: control task creation

Yandex Toloka

ProjectsUsersSkillsProfileMessages

?

\$0.00 / \$20.00

🇬🇧

Iron Man

Edit tasks

Use main tasks as a starting point to create control tasks or training tasks.  
Control tasks are for checking the quality of responses from performers. They contain correct responses to compare with actual responses.  
Training tasks are for teaching performers how to complete tasks. They contain correct responses and hints.  
[Learn more](#)

Main 100Control tasks 0Training tasks 0

Create control tasksCreate training tasksDownload

ID ↕	Overlap ↕	Responses from performers ↕	Last updated ↕
...18dd9fdc	3	0	08/04/2019 7:35:46 AM
...18dd9fcb	3	0	08/04/2019 7:35:46 AM
...18dd9fe1	3	0	08/04/2019 7:35:46 AM
...18dd9fc1	3	0	08/04/2019 7:35:46 AM
...18dd9fd0	3	0	08/04/2019 7:35:46 AM
...18dd9feb	3	0	08/04/2019 7:35:46 AM
...18dd9fde	3	0	08/04/2019 7:35:46 AM
...18dd9fc7	3	0	08/04/2019 7:35:46 AM



# Task: create a control task by answer selection

Yandex Toloka

ProjectsUsersSkillsProfileMessages

?

\$0.00 / \$20.00

🇬🇧

Iron Man

Projects > Are there shoes in the picture? > Are there shoes in the picture... > Uploaded tasks > Edit tasks

Control tasks: 15 

< >

Create control task

1. Specify the correct answers

2. Select the fields to include

✓

Field

Value

✓

result

Yes

Are there **shoes** in the picture?

☒ Yes

☐ No

☐ Picture not found

Save and go to next

ID	Details	Updated:
...1c9675c2 Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c967598 Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c9675c9 Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c9675a5 Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c96759e Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c9675ae Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c9675dd Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c9675cb Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c9675ba Main	Overlap: 3 Responses: 3	01/28/2020 16:33::
...1c9675cf Main	Overlap: 3 Responses: 3	01/28/2020 16:28::
...1c9675d1 Main	Overlap: 3 Responses: 3	01/28/2020 16:28::
...1c9675e0 Main	Overlap: 3 Responses: 3	01/28/2020 16:28::
...1c9675e2 Main	Overlap: 3 Responses: 3	01/28/2020 16:28::
...1c96759f	Overlap: 3	01/28/2020 16:28::

**Run the pool &  
result downloading**

# Pool running

Yandex Toloka

ProjectsUsersSkillsProfileMessages

?

\$0.00 / \$20.00

🇬🇧

Iron Man

Projects > Are there shoes in the picture? > First pool

First pool — closed

Statistics

Download results

▼

Edit

▼

?

POOL TASKS (File example for task uploading (tsv, UTF-8)) ?

UploadfilesDeleteEditPreview

~30task suites

90tasks

0training task

10control task

0%

Done 0

0

~ 30

POOL STATISTICS ?

0sec	—	(+)	\$	(+)	\$
------	---	-----	----	-----	----



# View and aggregate the results of your tasks

The screenshot displays the Yandex Toloka web interface for a requester. The browser address bar shows the URL `https://toloka.yandex.com/requester/project/23568/pool/6153926`. The page header includes the Yandex Toloka logo, navigation links (Projects, Users, Skills, Profile, Messages), and a user profile section for 'Iron Man' with a balance of \$0.00 / \$20.00.

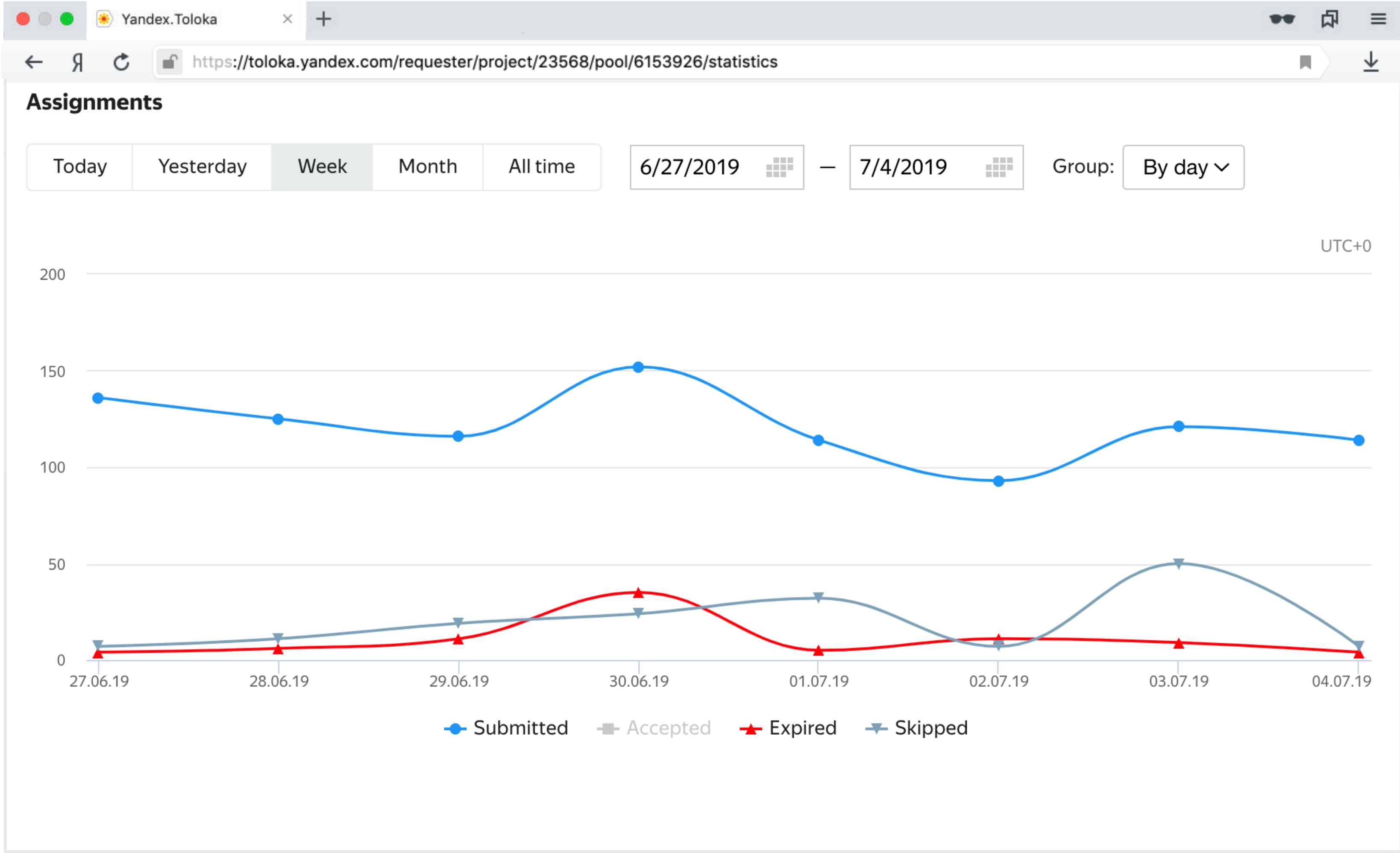
The main content area is titled 'First pool — closed' with a subtitle 'It's my first pool'. A red arrow points to the 'Download results' button in the top right corner of the pool section. A dropdown menu is open, showing options: 'View operations', 'Dawid-Skene aggregation model', and 'Aggregation by skill'. Another red arrow points to the 'View assignments' button located below the progress bar.

The 'POOL TASKS' section provides a summary of task statistics:

POOL TASKS (File example for task uploading (tsv, UTF-8))	
<b>30</b> task suites	<b>0</b> training task
<b>90</b> tasks	<b>10</b> control task

A progress bar at the bottom right indicates 100% completion, with the text 'Done 30, accepted 30' and a 'View assignments' button. The progress bar is a solid green bar from 0 to 30.

# Monitor the statistics on how tasks are performed



**Y**andex

**Thank you!  
Questions?**

**Evfrosiniya Zerminova**

Head of Data Analysis and Research Group



evfrosiniya-z@yandex-team.ru



<https://research.yandex.com/tutorials/crowd/wsdm-2020>