

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

Setting up and running label collection projects Instruction

Yandex Toloka for requesters:

https://toloka.ai/

For dataset:

http://tlk.s3.yandex.net/relsubstitutes/dataset_Y.tsv where Y is any number from 1 to 10

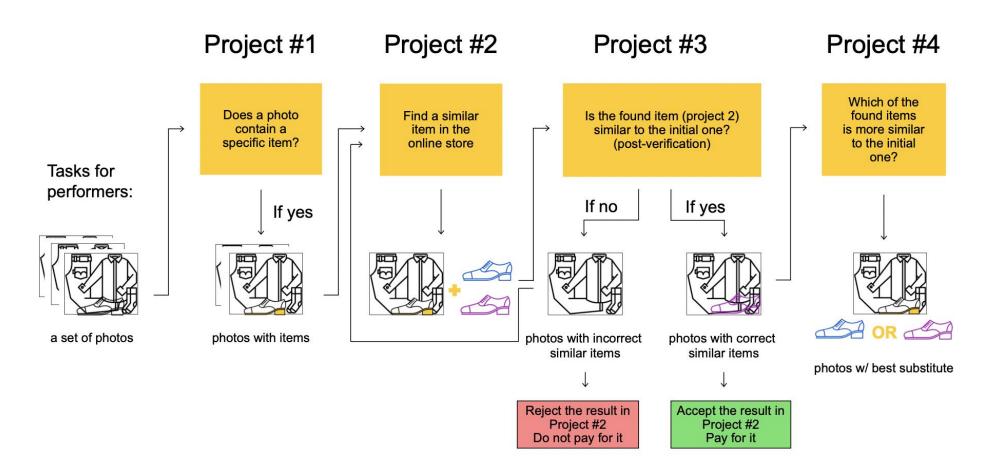
Tutorial slides:

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

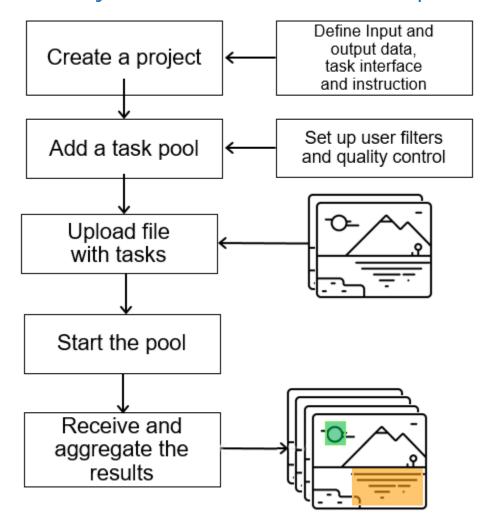
Table of contents

Suggested pipeline	3
Project creation. Main steps	4
Key types of instances in Yandex.Toloka	4
Project #1 Does the photo contain an item?	5
Project creation	6
Pool creation	9
Preparing and uploading a file with tasks	11
Receiving responses	14
Project #2 Find a similar item in an online store	15
Project creation	16
Pool creation	19
Preparing and uploading a file with tasks	23
Project #3 Does the item found look similar to the initial one?	24
Project creation (similar to the 1st project)	25
Pool creation	28
Preparing and uploading a file with tasks	30
Receiving responses	34
Upload reviewed results	35
Review assignments in the interface (another way of results validation)	36
Project #4 Which item is more similar?	37
Project creation	38
Pool creation	42
Preparing and uploading a file with tasks	44
Receiving responses	45
Annendia: Expanded code of the projects	16

Suggested pipeline



Project creation. Main steps

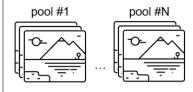


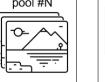
Key types of instances in Yandex. Toloka

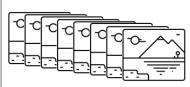
Project Pool

Task

- Defines the structure of tasks
- Defines how to perform them
- Is a batch of tasks
- Defines access of performers
- A particular input data
- Results for it from performers







Configure in a project:

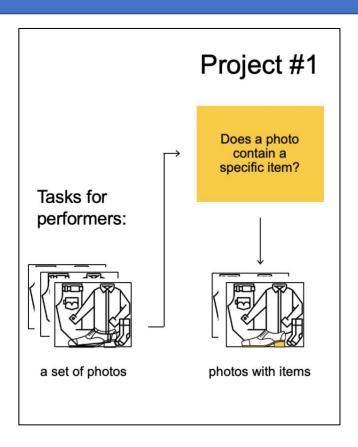
- Input and output data types
- Task interface
- Task instruction

Configure in a pool:

- Performer filters
- Quality control mechanisms
- Overlap settings



Project #1 Does the photo contain an item?



Important: Before you start using Toloka, make sure that the **English** language is selected.



Project creation

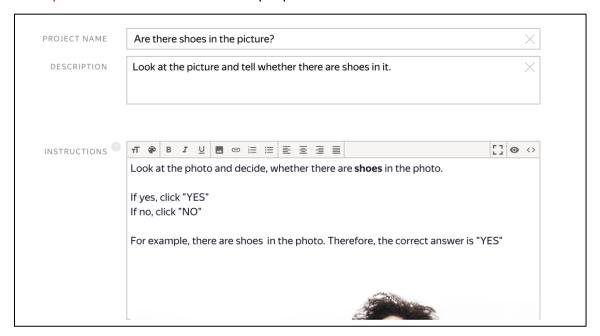
1. Click the button + Create project



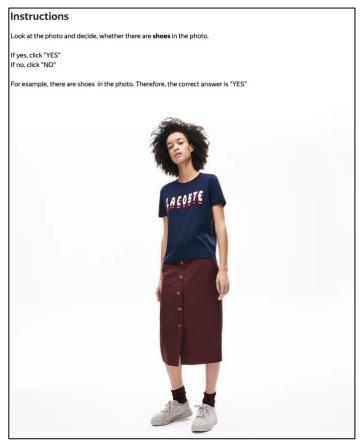
2. Choose the Image classification template.



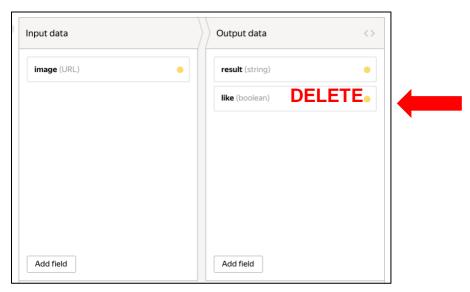
3. Enter a clear project name and description. Important: It will be visible for real people.



4. Write short and simple instructions. To include an image in the instruction just paste the link from the dataset provided by pressing button.



- 5. Define parameters for the input and output data:
 - The "image" input data field with the link type will be used to pass the image links to the performers.
 - You will be able to upload the file with links to the pool later.
 - The "result" field will be used to receive performer's responses.
 - The "like" field in the template is used to pass the response to the question "Do you like the photo?". Our project doesn't require this checkbox, so you don't need an output field for it. Let's remove it.



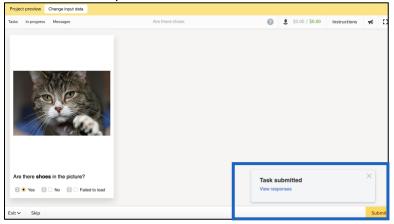
- 6. Create the task interface in the HTML block.
 - Delete the line with the checkbox component:

```
{{field type="checkbox" name="like" label="Do you like the photo?" hotkey="q"}}
```

- Replace "label" with your response options (change "Good" to "Yes" and "Bad" to "No"). Example:

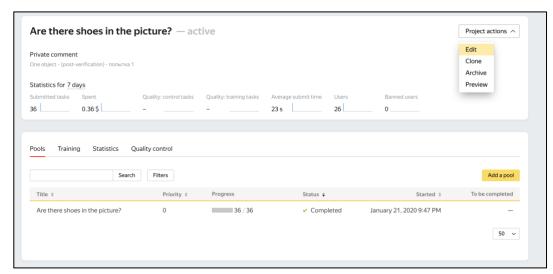
```
{{img src=image width="100%" height="400px"}}
 Are there <b>shoes</b> in the picture?
{{field type="radio" name="result" value="Yes" label="Yes"}}
{{field type="radio" name="result" value="No" label="No" }}
{{field type="radio" name="result" value="404" label="Picture not found" }}
```

- 7. Leave JavaScript and CSS block unchanged.
- 8. Click the **Preview** button to see the performer's view of the task. You will see tasks with standard pictures on the page. You will set the number of tasks per page when configuring a pool.
 - 9. Select the radio buttons in the preview and make sure that the task can be completed.



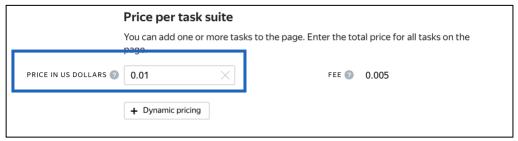
10. Click **Save** button to save the project.

Note. To edit project parameters, click the button in the list of projects or **Project actions** \rightarrow **Edit** on the project page.

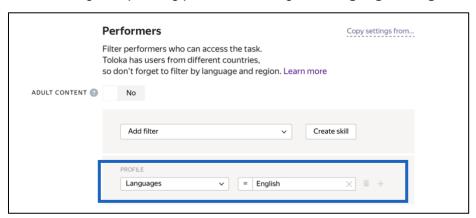


Pool creation

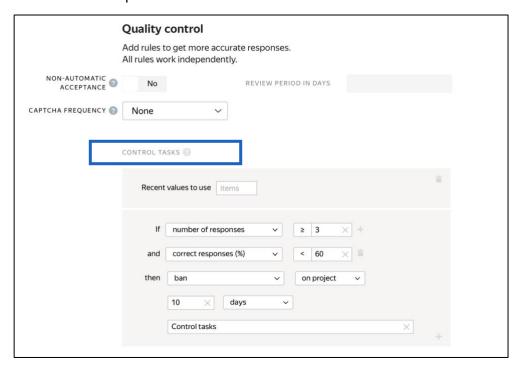
- 1. Click Add pool.
- 2. Give the pool any convenient name and description. You are the only one who can see them.
- 3. Specify the pool parameters:
 - Set the price per task page (for example, \$0.01).



- 4. Set up user filters.
 - Select English-speaking performers using the **Language = English** filter.



5. Set up quality control: Control tasks. Ban performers who give incorrect responses to control tasks. Example:



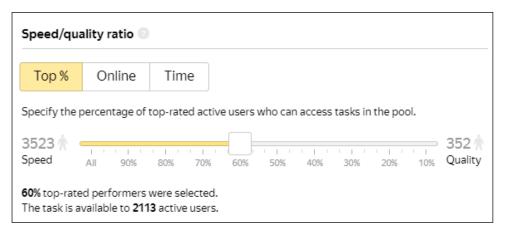
This rule will be triggered when the performer completes 3 control tasks in the pool. If the performer gives at least 3 responses to the control tasks and the percentage of correct responses is less than 60%, they lose access to the project for 10 days. If the percentage of correct responses is over 60%, the performer can pass to the next task page. The rule will be triggered after the next control task.

Optionally, add other quality control rules.

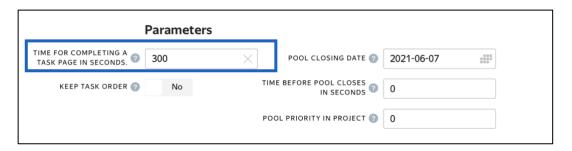
6. Overlap. This is the number of users who will complete the same task. For example, 3.



7. Optionally, specify the percentage of top-rated performers in the **Speed / Quality ratio**. **Important:** This can slow down pool completion.



8. Time allowed for completing a task page (for example, 300 seconds).



9. Save the pool.

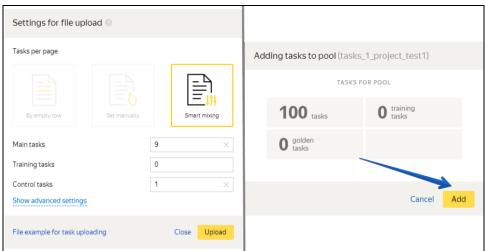
Preparing and uploading a file with tasks

- 1. Download TSV-file with images by link that you were provided at the beginning of the practice session.
- 2. Upload pool tasks from this file.

 Important: If you changed the name of the input field, change it in the file as well



Select **Smart mixing** and specify the number of tasks per page. For example: 9 main tasks and 1 control task.

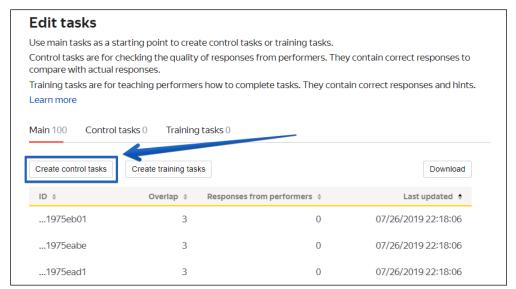


3. Create control tasks.

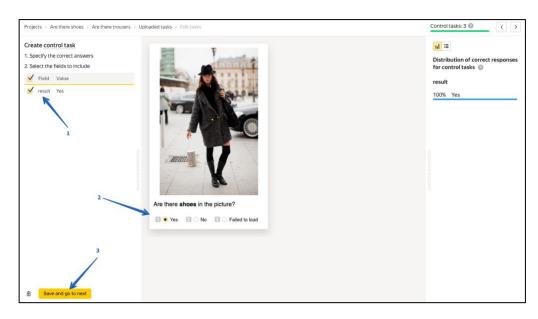
Note. Control tasks are tasks with the correct response known in advance. They are used to track the performer's quality of responses. The performer's response is compared to the response you provided. If they match, it means the performer answered correctly.

Click Edit → Create control tasks.



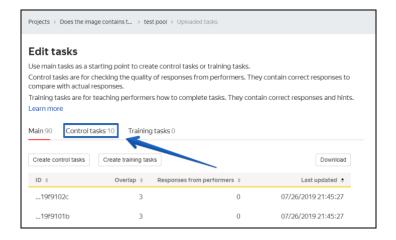


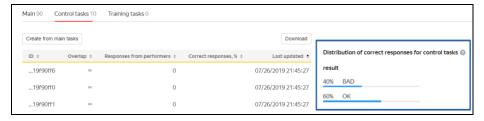
 Check the "result" output field that is used to match the user response to the control one, select the response and click Save and go to next.



Enter correct responses for your control tasks. In small pools, control tasks should account for 10-20% of all tasks.

Tip. Make sure to include different variations of correct responses in equal amounts. Open the **Control tasks** → **Distribution of correct responses for control tasks** tab.





Save the markup and check the number of control tasks.



4. Start the pool.

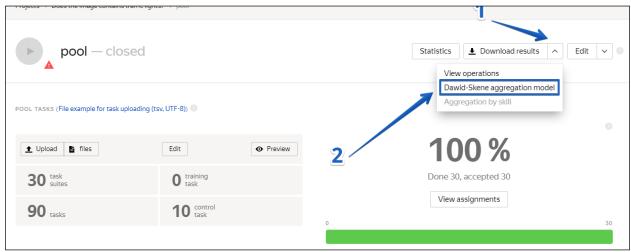
Important. Remember that real Toloka performers will complete the tasks. Double check that everything is correct with configuration of your project before you start the pool.



Receiving responses

Disclaimer: Aggregation takes from 5 to 20 minutes. During this time, you can start configuring your next project. Refresh the Operations page to check progress.

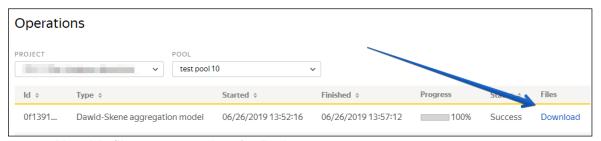
- 1. Wait until the pool is completed. Refresh the pool page to check progress.
- 2. Click the arrow next to the **Download results** button and run aggregation using the **Dawid-Skene model**.



 Go to the operations list and wait until aggregation finishes.
 Note. During this time, you can start working on your next project. Refresh the Operations page to check progress.

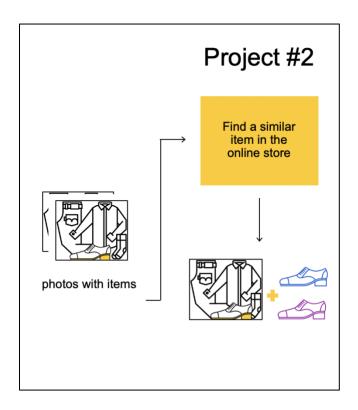


4. When aggregation is complete, download the TSV file with the results.

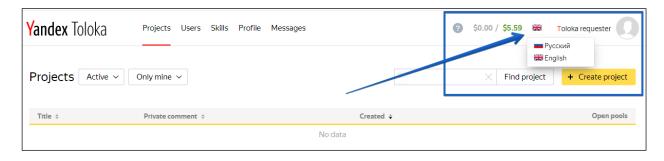


5. Use this file to prepare data for Project #2.

Project #2 Find a similar item in an online store



Important: If you just started using Toloka, make sure that the English language is selected.



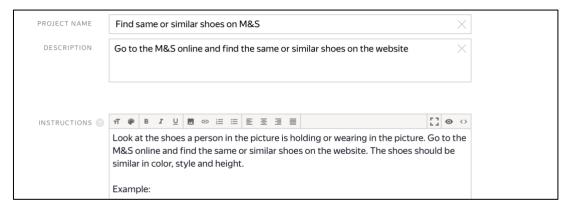
Project creation

Disclaimer: In this particular example we ask the performers to search for the clothing items on <u>Marks and Spencer</u>. However, you can use any other online store. In this case you must change Marks and Spencer to the store of your choice in Specifications, HTML and JS fields of the project.

1. Click the + Create project button and choose the Blank template.



2. Enter a clear project name and description. It will be shown to performers.



- 3. Write short and simple instructions.
- 4. Define specifications for the input and output data:

Input:

• The "**image**" input data field with the link type will be used to pass the image links to the performers.

Output:

- In the output field add "found_image", to give the performers space to upload the image.
- In the output field add "button" to check if the performer actually went to the required website
- In the output field add "found_link", here we will check whether the pattern of
 the url from the required website (a.g. online store, Marks and Spencer) matches
 the pattern of the submitted link. If you have chosen a different store then
 change "pattern": "https://www.marksandspencer.com/.*", to
 "pattern": "your_store.*"

The code for specifications are:

```
Input data:
  "image": {
    "type": "url",
    "hidden": false,
    "required": true
}
Output data:
  "button": {
    "type": "boolean",
    "hidden": false,
    "required": true,
    "allowed values": [
      true
    ]
  "found link": {
    "type": "string",
    "hidden": false,
    "pattern": "https://www.marksandspencer.com/.*",
    "required": true
  "found_image": {
    "type": "file",
    "hidden": false,
    "required": true
```

- 5. Create the task interface.
- Delete the whole HTML code in the template, and instead add the following code to show the initial item to the performers:

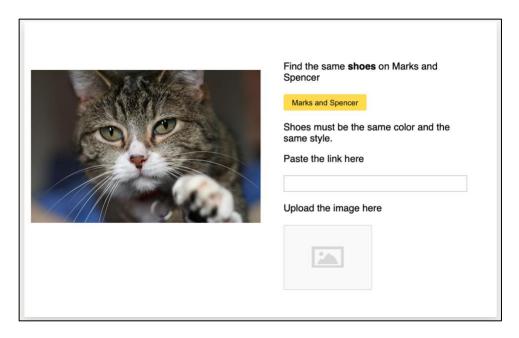
6. Now we need to check whether performer is going to submit a valid link and an image. To check it, DO NOT delete any of JS code

In case you are having trouble previewing call our team or check the expanded code in Appendix.

7. Add the following code in CSS field to set the images sizes proportionally:

```
.task {
  display: block;
  height: 500px;
  width: 800px;
}
.img {
  float: left;
  width: 50%;
}
.answers {
  float: left;
  width: 40%;
  margin: 5%;
}
```

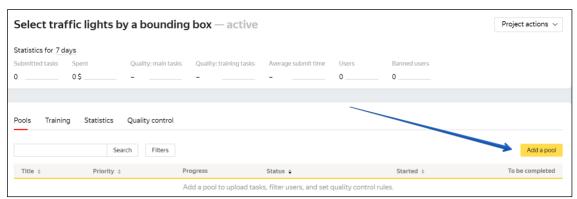
8. Click the **Preview** button to see the performer's view of the task. In this particular case you will not be able to submit the assignment in **Preview** as the image you will be trying to upload cannot be uploaded while the pool is still closed.



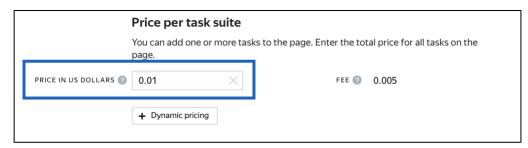
9. Click **Save** button to save the project.

Pool creation

1. Click Add pool.



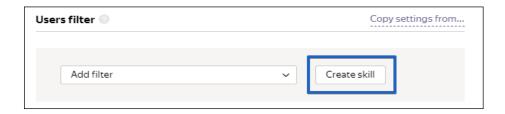
- 2. Give the pool any convenient name and description. You are the only one who can see them.
- 3. Specify the pool parameters:
 - Price per task page (for example, \$0.01)



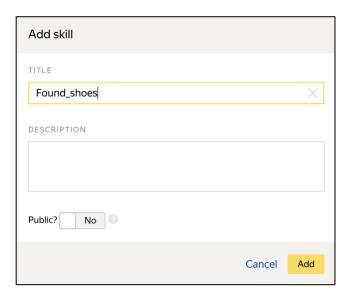
- 4. Set up user filters.
 - Select English-speaking performers using the Language = English filter.



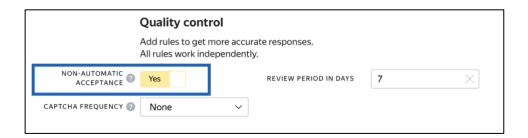
Create the "Found_shoes" skill that will be assigned to users after they
complete the pool tasks. You will use this skill to prevent these users from
checking tasks in the next project. Click Create skill:



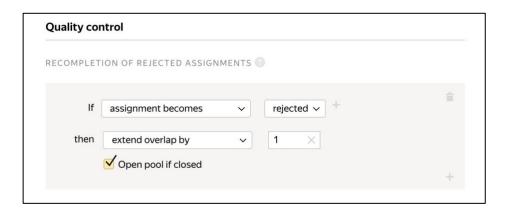
• Enter the skill name and add a description if desired. You are the only one who will see it. Leave the skill **private**, as it is by default and click **Add.**



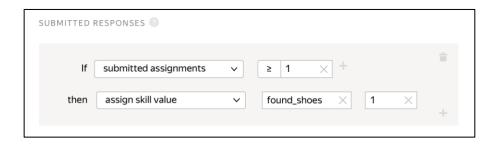
5. Turn on the **Non-automatic acceptance** option and enter the number of days for checking in the **Deadline** field (for example, 7).



- 6. Set up quality control:
 - Resend the rejected tasks for completion. Add the <u>Recompletion of rejected</u> assignments:

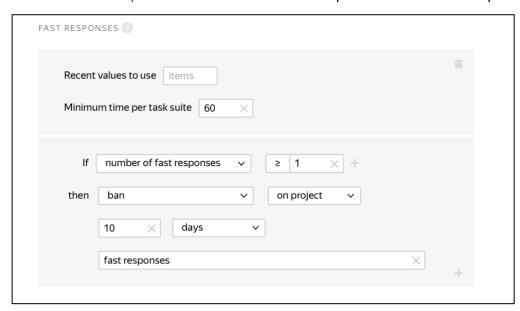


• <u>Submitted responses</u>. Add a rule to mark users who completed at least one task in the pool.

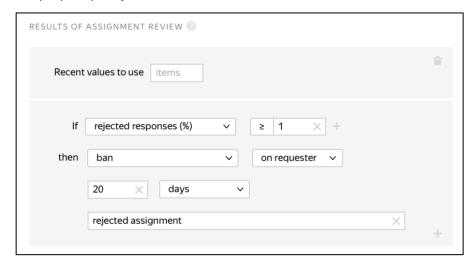


Tip. If the skill you created doesn't appear in the drop-down list, save the pool, and then open it for editing again.

Add <u>Fast responses</u> rule to block those who provide information suspiciously fast



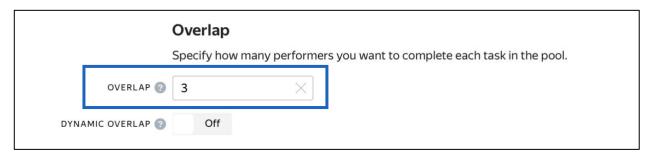
 Add <u>Results of assignments review</u> rule to ban those who brought results of improper quality



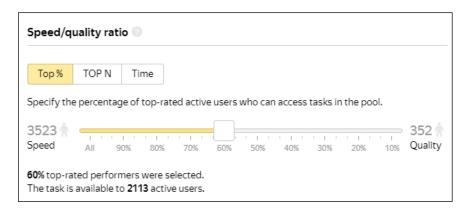
Optionally, add other quality control rules.

Tip. Control tasks and majority vote are not used in this type of project, because performer's links and photos that she will provide must exactly match the reference, which is practically impossible.

7. Overlap. This is the number of users who will complete the same task. Because we want various options for each photo, put overlap equal to 3.

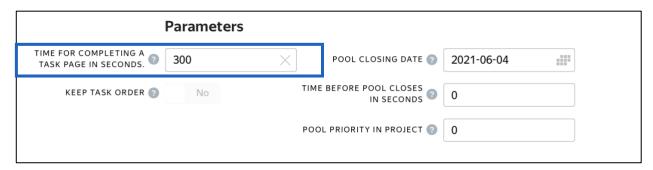


8. Optionally, specify the percentage of top-rated performers in the **Speed / Quality ratio**.



Important: This can slow down pool completion.

9. Time allowed for completing a task page (for example, 300 seconds)



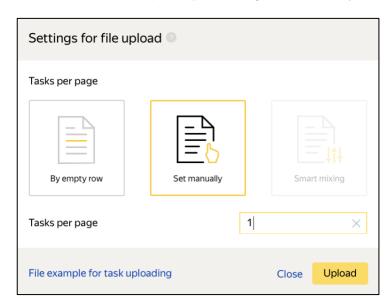
10. Save the pool.

Preparing and uploading a file with tasks

- 1. Open the file with aggregated results from the project #1.
- 2. Select only images suitable for highlighting (**OK** answers or another value if you have changed it in the **"result"** field). Use a text editor or a spreadsheet editor.
- 3. Copy the column with the selected links to a new page or document and give a name to the **INPUT:image** column.

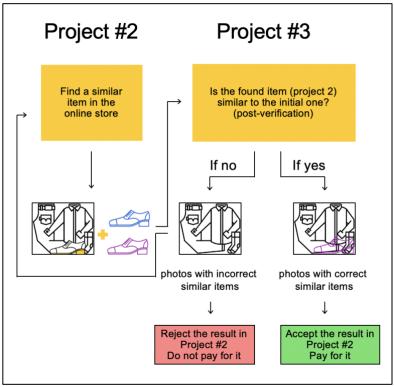
Important: If you changed the input field name in the project to something other than "image", change the name in the file as well: INPUT:<your field name>.

- 4. Save the file in TSV format.
- 5. Upload the file to the pool by selecting **Set manually**. Set 1 task per page.



6. Start the pool.

Project #3 Does the item found look similar to the initial one?



Important: If you just started using Toloka, make sure that the **English** language is selected.

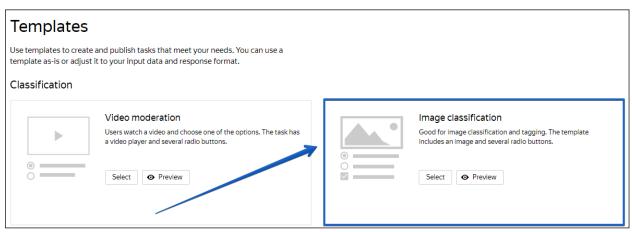


Project creation (similar to the 1st project)

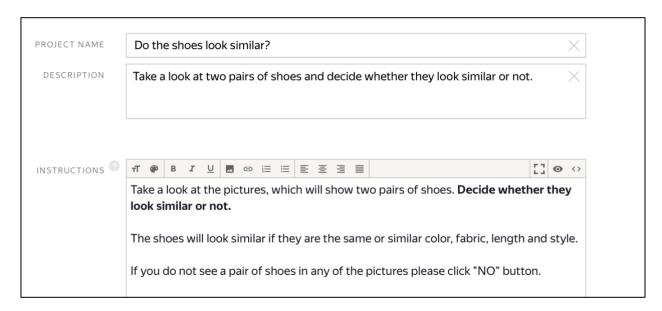
1. Click the button + Create project



2. Choose the Image classification template.



3. Enter a clear project name and description. Important: It will be visible for real people.



- 4. Write short and simple instructions.
- 5. Define parameters for the input and output data:

Input:

- The "image" input data field with the "url" type will be used to pass the initial image links to the performers.
- The "found_link" field with the "url" type will allow performers can go to the website.
- The "assignment_id" field with the "string" type, will be used to pass the number of the completed task.

Output:

Leave "result" as it is.

The code for specifications is:

```
Input data:
  "image": {
    "type": "url",
    "hidden": false,
    "required": true
  "found link": {
    "type": "url",
"hidden": false,
    "required": true
  "assignment_id": {
    "type": "string",
    "hidden": true,
    "required": true
}
Output data:
  "result": {
    "type": "string",
    "hidden": false,
    "required": true
}
```

6. Delete the whole HTML code in the template, and add the following one.

```
{{img src=image height="400px"}}
{{iframe src=found_link height="600px"}}

Check that the uploaded image matches the product in the store.
{{button label="Check the item" href=found_link action=true}}

Are these <b>shoes</b> similar to each other?
Shoes must be the same color and the same style.
{{field type="radio" name="result" value="Yes" label="Yes"}}
{{field type="radio" name="result" value="No" label="No"}}
</div>
```

- 7. Leave the JS block unchanged.
- 8. In the CSS block paste the following (don't forget the other half on the next page):

```
.task {
display: block;
min-height: 620px;
width: 100%;
box-sizing: border-box;
width: calc(100% - 100px);
}
.img {
float: left;
width: 30%;
}
```

```
.iframe {
float: left;
width: 48%;
margin-left: 10px;
}
.text {
float: left;
width: 18%;
margin-left: 10px;
}
```

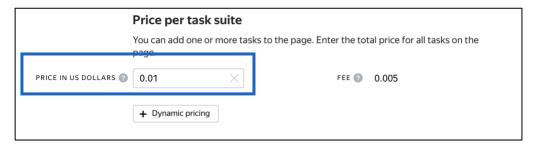
9. Click the **Preview** button to see the performer's view of the task.



- 10. Select the radio buttons in the preview and make sure that the task can be completed.
- 11. Click Save button to save the project.
 Note. To edit project parameters, click the button in the list of projects or Project actions → Edit on the project page.

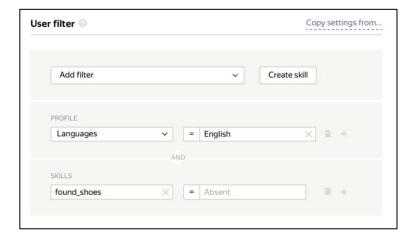
Pool creation

- 1. Click Add pool.
- 2. Give the pool any convenient name and description. You are the only one who can see them.
- 3. Specify the pool parameters:
 - Set the price per task page (for example, \$0.01).

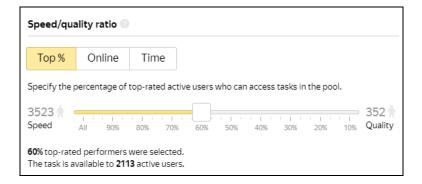


- 4. Set up user filters.
 - Select English-speaking performers using the "Language = English" filter.
 Prevent performers who completed previous tasks from checking this one. To do this, set a filter with the <u>"Found_shoes" skill</u>:

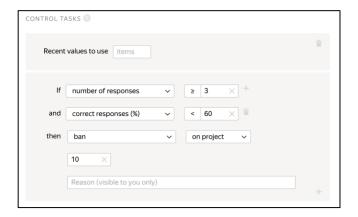
The **"Found_shoes"** skill = absent (empty field)



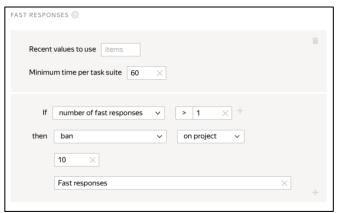
Optionally, specify the percentage of top-rated performers in the **Speed / Quality ratio**. **Important:** This can slow down pool completion.



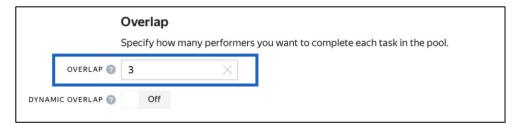
5. Set up <u>quality control</u>: <u>Golden Set aka Control tasks</u>. Ban performers who give incorrect responses to control tasks. Example:



<u>Fast responses</u>. You can ban the performers who suspiciously fast responses. This way you can get rid of cheaters in your pool. Example:



6. Overlap. This is the number of users who will complete the same task. For example, 3 is enough for aggregation.



- 7. Time allowed for completing a task page (for example, 300 seconds).
- 8. Keep task order for your convenience



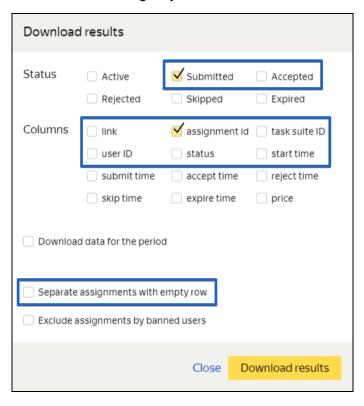
9. Save the pool.

Preparing and uploading a file with tasks

- 1. Wait until the pool of project #2 on "finding similar shoes" is completed.
- 2. Open the pool page in Project #2 and click the **Download results** button



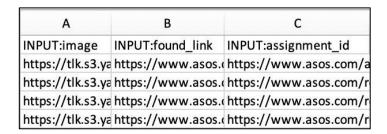
- Clear the **Accepted** checkbox and select **Submitted**.
- Clear link, user ID, status, start time and Separate assignments with empty row checkboxes. This will give you a list of unreviewed tasks.



- 3. Keep and rename the following columns:
- Keep the name of the "INPUT:image" column as it is.

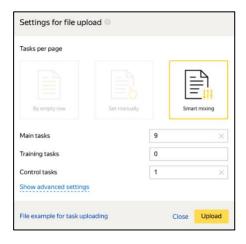
- Change the name of the "OUTPUT:found_link" column to "INPUT:found_link". To check this image for correctness in project 3.
- Change the "ASSIGNMENT:assignment_id" column name to INPUT:assignment_id" to later track and match the assignment number.

Make sure you have the headers of the columns exactly as below



Save the file in TSV format.

- 4. Open the pool page in Project #3.
- 5. Upload the file to the pool by selecting **Smart mixing**. Set the number of main and control tasks per page (for example, 9 and 1).



6. Create control tasks.

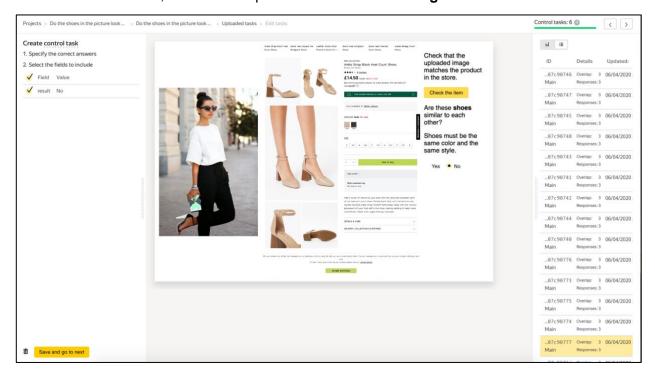
Note. Control tasks are tasks with the correct response known in advance. They are used to track the performer's quality of responses. The performer's response is compared to the response you provided. If they match, it means the performer answered correctly.

Click Edit → Create control tasks.



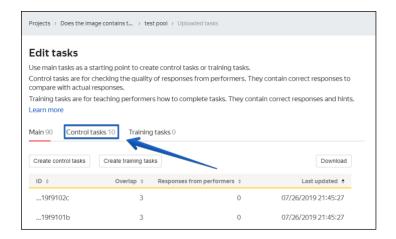


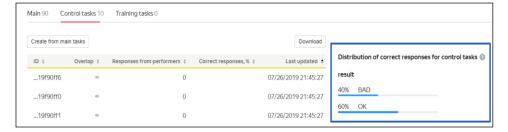
 Check the "result" output field that is used to match the user response to the control one, select the response and click Save and go to next.



Enter correct responses for your control tasks. In small pools, control tasks should account for approximately 10% of all tasks.

Tip. Make sure to include different variations of correct responses in equal amounts. Open the **Control tasks** → **Distribution of correct responses for control tasks** tab.





Save the markup and check the number of control tasks.



7. Start the pool.

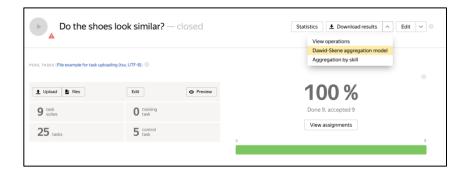
Important. Remember that real Toloka performers will complete the tasks. Double check that everything is correct with configuration of your project before you start the pool.



Receiving responses

Disclaimer: Aggregation takes from 5 to 20 minutes. During this time, you can start configuring your next project. Refresh the Operations page to check progress.

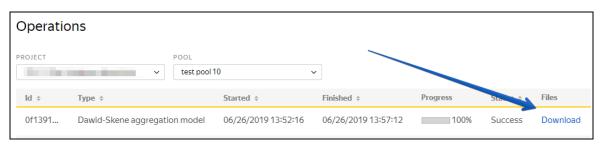
- 1. Wait until the pool is completed.
- 2. Click the arrow next to the **Download results** button and run aggregation using the **Dawid-Skene model**.



3. Go to the operations list and wait until aggregation finishes.



4. Download the responses.



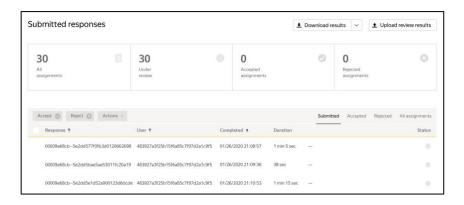
Upload reviewed results

As you set **post verification** in the <u>pool settings in Project #2</u>, you need to check the performers' responses within the time limit set in the **Deadline** field.

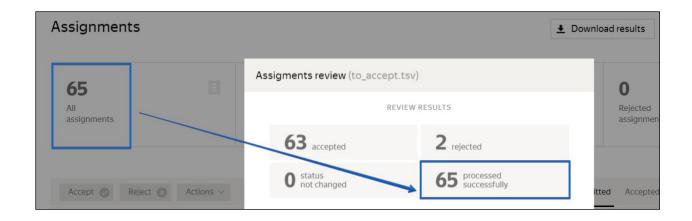
- 1. Open the file with aggregated results in a spreadsheet editor.
- 2. Add the following columns:
 - "ACCEPT:verdict" The result of verification.
 - "ACCEPT:comment" Comments for performers if responses were rejected (for example, which part of the instructions wasn't followed).
- 3. Change the name of the "INPUT:assignment_id" column to "ASSIGNMENT:assignment_id".
- 4. Delete all other columns
- 5. Fill in the "ACCEPT:verdict" and "ACCEPT:comment" columns:
 - If the aggregated result for the task is OK, put "+" then the task will be accepted.
 - If the result is BAD or 404, put "-" then the task will be rejected. Enter the reason for rejection in the "ACCEPT:comment" field . For example: The item provided is incorrect or improper.
- 6. Now you can delete the other columns. Save the edited TSV file.

А	В	С
ASSIGNMENT:assignment_id	ACCEPT:verdict	ACCEPT:comment
00009c42455e27381d139baa	+	
00009c42455e27337bb0e86f0	-	The item provided is incorrect
00009c42455e2737692f3cb80	+	
00009c42455e2739bcb0e86f0	+	

- 7. Open the pool page in Project #2.
- 8. Click Review assignments on the pool page above the progress bar.
- 9. Click Upload review results.



- 10. Select the file and upload it to Toloka.
- 11. Check that all tasks have changed their status to accepted or rejected.



12. You rejected tasks and set up the rule to send them for re-completion while configuring Project 2.

If have enough time you can do as many more reiterations as needed in order to receive as much clean data as you can. The steps are the following: The pool will open again, and these tasks will be resent to other performers. After the pool is marked up, download the new results and submit them for review. Download the reviewed results. Repeat these steps until all the images from the second project are correctly marked up.

But if you do not have enough time, move on to the next project and you can complete the reiterations later at your own pace.

Review assignments in the interface (another way of results validation)

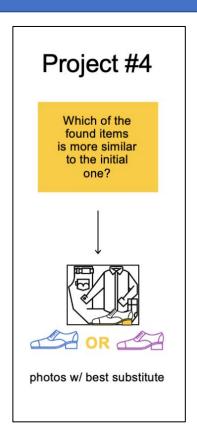
You can also <u>review assignments by yourself</u> and see the results of the crowdsourcing pipeline that you have created.

- 1. Open the pool page in Project #2.
- 2. Click the **Review assignments** button on the pool page.

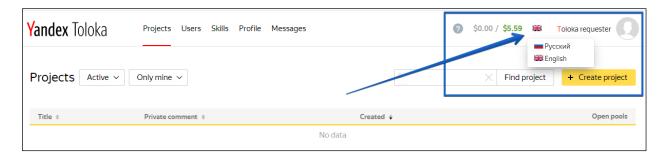


- Choose an assignment then click Accept or Reject.
- For rejected assignments, enter a comment (explain why you decline it).

Project #4 Which item is more similar?



Important: Before you start using Tolok a, make sure that the **English** language is selected.

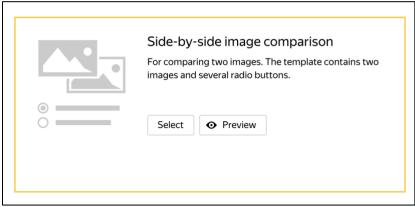


Project creation

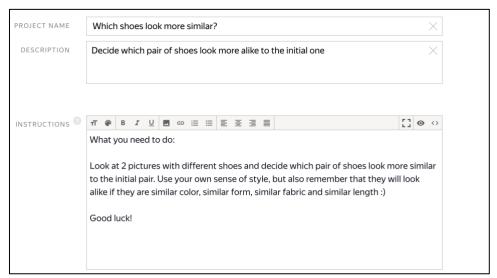
1. Click the button + Create project



2. Choose the **Side-by-side image comparison** template.



- Enter a clear project name and description.
 Important: It will be visible for real Toloka performers.
- 2. Write short and simple instructions. Example:



- 11. Define parameters for the input and output data:
 - The "image" field is the initial image from the dataset.
 - The "**left_link**" field is a link that the performer provided to match the item from the initial image.
 - The "right_link" field is another link the performer provided.

You will be able to upload the file with links to the pool later.

The "result" field will be used to receive performer's responses.

The code for specifications is: Input data:

```
"image": {
    "type": "url",
    "hidden": false,
    "required": true
  "left_link": {
    "type": "url",
    "hidden": false,
    "required": true
  "right link": {
    "type": "url",
    "hidden": false,
    "required": true
 }
}
Output data:
  "result": {
    "type": "url",
    "hidden": false,
    "required": true
  }
}
```

12. Create the task interface in the HTML block.

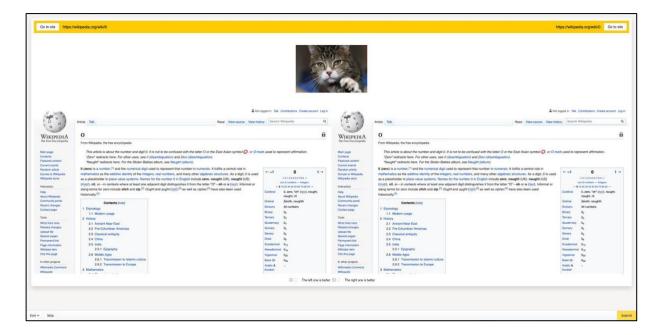
```
<div class="header">
    <div class="left caption">
       {{button label="Go to site" href=uploaded link left size="L"}}
       {{uploaded link left}}
   </div>
    <div class="right caption">
       {{uploaded link right}}
       {{button label="Go to site" href=uploaded link right size="L"}}
   </div>
</div>
{{img src=image}}
<div class="content clearfix">
   <div class="left page">
      {{iframe src=uploaded_link_left width="100%" height="700px" real-
size=true screenshot=true}}
   </div>
   <div class="right page">
       {{iframe src=uploaded link right width="100%" height="700px" real-
size=true screenshot=true}}
   </div>
</div>
<div class="footer">
   {{field type="radio" name="result" label="The left one is better"
value=result_left hotkey="1"}}
   {{field type="radio" name="result" label="The right one is better"
value=result_right hotkey="2"}}
</div>
```

13. DO NOT delete any of JS code but add the following JS block right before OnRender

```
getTemplateData: function() {
           var data = TolokaHandlebarsTask.prototype.getTemplateData.apply(this,
   arguments),
           input = this.getTask().input_values;
           var left link = input.left link;
           var right link = input.right link;
           var uploaded link left = '',
               uploaded link right = ''
           if (Math.floor(Math.random() * 2)) {
               uploaded link left = left link
               uploaded link right = right link
           } else {
               uploaded_link_left = right_link
               uploaded_link_right = left_link
          data.uploaded_link_left = uploaded_link_left;
           data.uploaded_link_right = uploaded_link_right;
           data.result_left = uploaded_link left;
           data.result_right = uploaded_link_right;
           return data;
       },
14. In the CSS block add:
   .task {
     display: block;
     text-align:center;
   }
   .header {
     overflow: hidden;
     background-color: #FFCC00;
   .caption {
     width: 50%;
   .url {
    white-space: nowrap;
    overflow: hidden;
     text-overflow: ellipsis;
    max-width: calc(100% - 182px);
     display: inline-block;
     vertical-align: bottom;
   }
   .button {
    margin: 10px;
    max-width: 182px;
   }
   .button__label {
    white-space: nowrap;
     overflow: hidden;
     text-overflow: ellipsis;
     max-width: 150px;
   .content {
```

```
margin: 10px 0;
}
.page {
    display: inline-block;
    width: 50%;
}
.left {
    float: left;
    text-align: left;
}
.right {
    float: right;
    text-align: right;
}
.clearfix {
    overflow: hidden;
    width: 100%;
}
```

15. Click the **Preview** button to see the performer's view of the task. You will see standard pictures on the page.

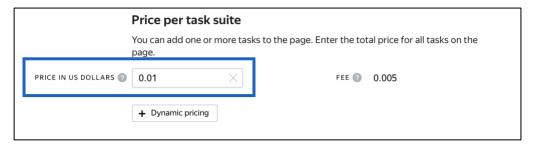


- 16. Select the radio buttons in the preview and make sure that the task can be completed.
- 17. Click **Save** button to save the project.

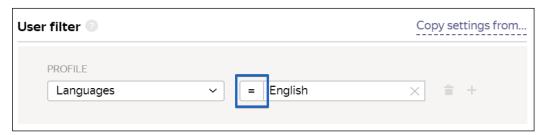
Note. To edit project parameters, click the button in the list of projects or **Project** actions \rightarrow Edit on the project page.

Pool creation

- 1. Click Add pool.
- 2. Give the pool any convenient name and description. You are the only one who can see them.
- 3. Specify the pool parameters:
 - Price per task page (for example, \$0.01)

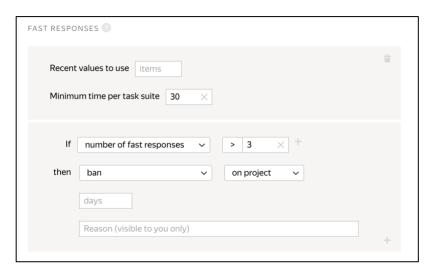


- 4. Set up user filters.
 - Select English-speaking performers using the Language = English filter.



5. Set up quality control:

<u>Fast responses</u>. You can ban the performers who suspiciously fast responses. This way you can get rid of cheaters in your pool. Example:



Optionally, add other quality control rules.

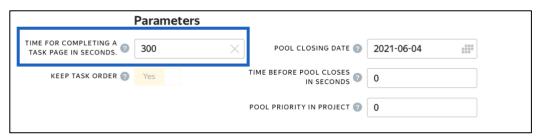
6. Overlap. This is the number of users who will complete the same task. Put a larger number in this task. For example, 10.



7. Optionally, specify the percentage of top-rated performers in the **Speed / Quality ratio**. **Important:** This can slow down pool completion.



8. Time allowed for completing a task page (for example, 300 seconds).



9. Save the pool.

Preparing and uploading a file with tasks

- 1. Take the downloaded TSV file with validated responses from Project 3.
- 2. Now you need to generate pairs for each INPUT:image so that you will be able to compare two found images with the initial one and decide which one is more similar than another.

You can either generate the pairs by hand, using MS Excel or any other editor or you can automatically do it. We recommend using Python and Jupyter Lab.

You can consult with our results https://tlk.s3.yandex.net/wsdm2020/SbS_Toloka_prep&aggr_data.ipynb

3. <u>Upload pool tasks</u> from this file. <u>Important:</u> If you changed the name of the input field, change it in the file as well



<u>Upload the file</u> to the pool by selecting **Set manually** and specify the number of tasks per page.

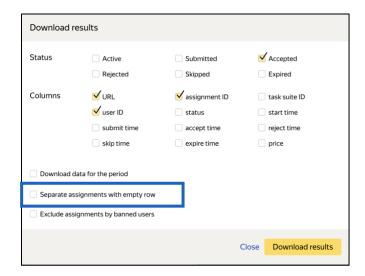
You can experiment with the number of tasks!

4. Start the pool.

Important. Remember that real Toloka performers will complete the tasks. Double check that everything is correct with configuration of your project before you start the pool.

Receiving responses

- 1. Wait until the pool is completed. Refresh the pool page to check progress.
- 2. Download **the accepted** results. Select the URLs, user IDs and assignment IDs like in the picture below. Do not forget to clear "Separate assignments with empty row" box!



3. Try to run Bradley Terry model on these results (you can consult with our results https://tlk.s3.yandex.net/wsdm2020/SbS_Toloka_prep&aggr_data.ipynb)

GOOD LUCK!!!

Appendix: Expanded code of the projects

Project 1

Specifications:

Input:

```
"image": {
    "type": "url",
    "hidden": false,
    "required": true
}
```

Output:

```
{
  "result": {
    "type": "string",
    "hidden": false,
    "required": true
}
}
```

HTML:

```
{{img src=image width="100%" height="400px"}}  
<div>Are there <b>shoes</b> in the picture?</div>  
<div> {{field type="radio" name="result" value="Yes" label="Yes" hotkey="1"}}  
{{field type="radio" name="result" value="No" label="No" hotkey="2"}}</div>
```

JS:

```
exports.Task = extend(TolokaHandlebarsTask, function (options) {
   TolokaHandlebarsTask.call(this, options);
}, {
   onRender: function() {
        // DOM element for task is formed (available via #getDOMElement())
},
   onDestroy: function() {
        // Task is completed. Global resources can be released (if used)
}
});

function extend(ParentClass, constructorFunction, prototypeHash) {
   constructorFunction = constructorFunction || function () {};
   prototypeHash = prototypeHash || {};
   if (ParentClass) {
      constructorFunction.prototype = Object.create(ParentClass.prototype);
   }
   for (var i in prototypeHash) {
```

```
constructorFunction.prototype[i] = prototypeHash[i];
}
return constructorFunction;
}
```

Project 2

Specifications:

Input:

```
{
  "image": {
    "type": "url",
    "hidden": false,
    "required": true
  }
}
```

Output:

```
"button": {
   "type": "boolean",
   "hidden": false,
    "required": true,
    "allowed values": [
    true
 "found link": {
   "type": "string",
    "hidden": false,
    "pattern":
"https://www.marksandspencer.com/.*",
    "required": true
 "found image": {
   "type": "file",
   "hidden": false,
   "required": true
 }
```

HTML:

JS:

```
exports.Task = extend(TolokaHandlebarsTask, function (options) {
 TolokaHandlebarsTask.call(this, options);
}, {
 onRender: function() {
   // DOM element for task is formed (available via #getDOMElement())
  onDestroy: function() {
  // Task is completed. Global resources can be released (if used)
});
function extend(ParentClass, constructorFunction, prototypeHash) {
 constructorFunction = constructorFunction || function () {};
 prototypeHash = prototypeHash || {};
 if (ParentClass) {
   constructorFunction.prototype = Object.create(ParentClass.prototype);
  for (var i in prototypeHash) {
  constructorFunction.prototype[i] = prototypeHash[i];
 return constructorFunction;
CSS:
.task {
 display: block;
 height: 500px;
width: 800px;
.img {
 float: left;
 width: 50%;
.answers {
 float: left;
 width: 40%;
margin: 5%;
```

Project 3

Input:

Specifications:

```
"image": {
    "type": "url",
    "hidden": false,
    "required": true
},
"found_link": {
    "type": "url",
    "hidden": false,
    "required": true
},
```

Output:

```
{
  "result": {
    "type": "string",
    "hidden": false,
    "required": true
}
```

```
"assignment_id": {
    "type": "string",
    "hidden": true,
    "required": true
}
```

HTML:

```
{{img src=image height="400px"}} {{iframe src= found link height="600px"}}
Check that the uploaded image matches the product in the store.
{{button label="Check the item" href=found link action=true}}
Are these <b>shoes</b> similar to each other?
Shoes must be the same color and the same style.
{{field type="radio" name="result" value="Yes" label="Yes"}}
{{field type="radio" name="result" value="No" label="No"}}
JS:
exports.Task = extend(TolokaHandlebarsTask, function (options) {
  TolokaHandlebarsTask.call(this, options);
}, {
  onRender: function() {
  // DOM element for task is formed (available via #getDOMElement())
  onDestroy: function() {
  // Task is completed. Global resources can be released (if used)
});
function extend(ParentClass, constructorFunction, prototypeHash) {
 constructorFunction = constructorFunction || function () {};
  prototypeHash = prototypeHash || {};
  if (ParentClass) {
   constructorFunction.prototype = Object.create(ParentClass.prototype);
  for (var i in prototypeHash) {
  constructorFunction.prototype[i] = prototypeHash[i];
  return constructorFunction;
CSS:
.task {
display: block;
min-height: 620px;
width: 100%;
box-sizing: border-box;
width: calc(100% - 100px);
}
```

```
.img {
float: left;
width: 30%;
}
.iframe {
float: left;
width: 48%;
margin-left: 10px;
}
.text {
float: left;
width: 18%;
margin-left: 10px;
}
```

Project 4:

Specifications:

```
Input:
```

```
{
    "image": {
        "type": "url",
        "hidden": false,
        "required": true
},
    "left_link": {
        "type": "url",
        "hidden": false,
        "required": true
},
    "right_link": {
        "type": "url",
        "hidden": false,
        "required": true
}
```

Output:

```
{
  "result": {
    "type": "url",
    "hidden": false,
    "required": true
}
```

HTML:

```
{{iframe src=uploaded link left width="100%" height="700px"
real-size=true screenshot=true}}
     </div>
     <div class="right page">
         {{iframe src=uploaded link right width="100%" height="700px"
real-size=true screenshot=true}}
     </div>
</div>
<div class="footer">
    {{field type="radio" name="result" label="The left one is better"
value=result left hotkey="1"}}
    {{field type="radio" name="result" label="The right one is better"
value=result right hotkey="2"}}
</div>
JS:
exports.Task = extend(TolokaHandlebarsTask, function(options) {
   TolokaHandlebarsTask.call(this, options);
   getTemplateData: function() {
       var data =
TolokaHandlebarsTask.prototype.getTemplateData.apply(this, arguments),
           input = this.getTask().input values;
       var left link = input.left link;
       var right link = input.right link;
       var uploaded link left = '',
           uploaded link right = ''
        if (Math.floor(Math.random() * 2)) {
            uploaded_link_left = left_link
            uploaded link right = right link
        } else {
           uploaded link left = right link
           uploaded link right = left link
        data.uploaded link left = uploaded link left;
        data.uploaded link right = uploaded link right;
        data.result left = uploaded_link_left;
       data.result right = uploaded link right;
 return data;
},
   onRender: function() {
    // DOM element for task is formed (available via #getDOMElement())
   },
   onDestroy: function() {
    // Task is completed. Global resources can be released (if used)
   }
});
function extend(ParentClass, constructorFunction, prototypeHash) {
   constructorFunction = constructorFunction || function() {};
   prototypeHash = prototypeHash || {};
   if (ParentClass) {
       constructorFunction.prototype =
Object.create(ParentClass.prototype);
for (var i in prototypeHash) {
```

```
constructorFunction.prototype[i] = prototypeHash[i];
  return constructorFunction;
CSS:
.task {
display: block;
text-align:center;
.header {
 overflow: hidden;
background-color: #FFCC00;
.caption {
width: 50%;
.url {
white-space: nowrap;
overflow: hidden;
text-overflow: ellipsis;
max-width: calc(100% - 182px);
display: inline-block;
vertical-align: bottom;
.button {
margin: 10px;
max-width: 182px;
}
.content {
margin: 10px 0;
.page {
display: inline-block;
 width: 50%;
.left {
float: left;
text-align: left;
}
.right {
float: right;
text-align: right;
.clearfix {
overflow: hidden;
```

```
width: 100%;
}
.image {
  display: inline-block;
  width: 50%;
}
```