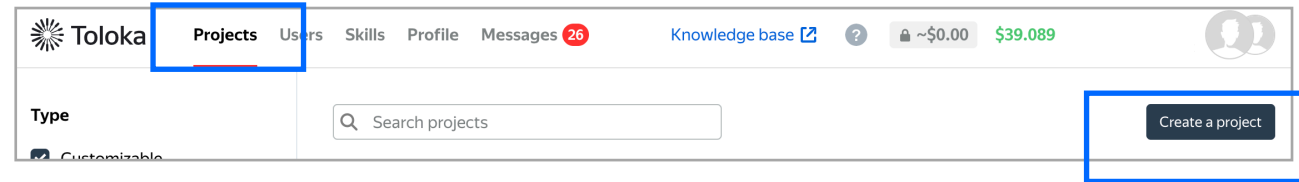# Toloka

manual

Welcome to Toloka!

Now you can create tasks and mark up data. To launch your first project in Toloka, just complete the following steps:

✓ Create your first project

✓ Create and setup a task pool and upload data

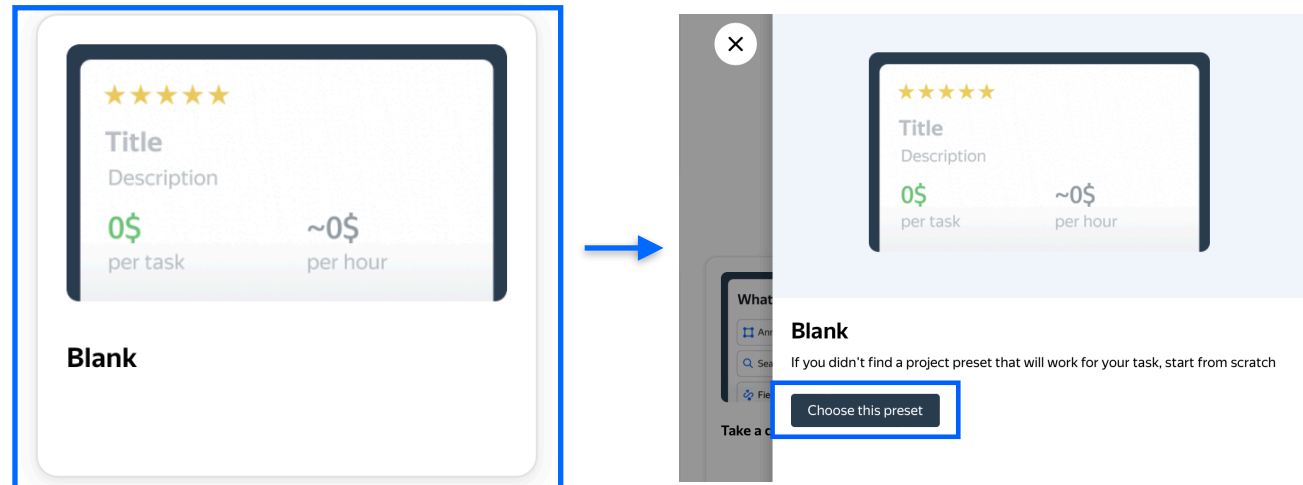✓ Get results

Create a project

Toloka

Project creation

1. Go to **Projects** and Click the button **+ Create project**

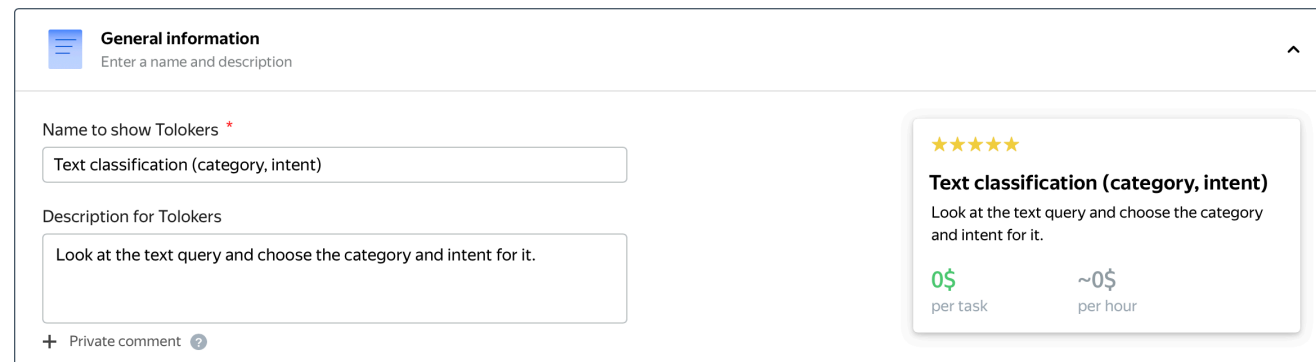2. Choose the **blank** template and click the button + **Choose this preset**

3. Enter a clear project name and description in General information

   Important: It will be visible to others

Don't see a good match?

New project

General information
Enter a name and description

Name to show Tolokers *

Text classification (category, intent)

Description for Tolokers

Look at the text query and choose the category and intent for it.

+ Private comment

★★★★★
Text classification (category, intent)
Look at the text query and choose the category and intent for it.

0$ per task   ~0$ per hour

4. Update the task interface in the **Template Builder** block

Task interface
Create a user-friendly interface for performers

Editor

○ HTML / JS / CSS ?   ● Template builder ?

4.1. Delete all code you see there and replace it on [another](#) one in **Config** section


```
Config

1   {
2       "view": {
3           "items": [
4               {
5                   "content": {
6                       "path": "query",
7                       "type": "data.input"
8                   },
9                   "label": "Query",
10                  "type": "view.text"
11              },
12              {
13                  "data": {
14                      "path": "domain",
15                      "type": "data.output"
16                  },
17                  "options": [
18                      {
19                          "value": "kitchen",
20                          "label": "Kitchen"
21                      },
22                      {
23                          "value": "travel",
24                          "label": "Travel"
25                      }
26                  ],
27                  "label": "Please select query category:",
28                  "validation": {
29                      "type": "condition.required"
```
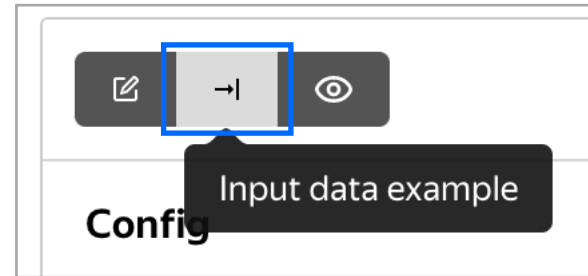
Preview

Query

Please select query category:
Q Kitchen   W Travel

4.2. Now you can change input data to see how it will be displayed for annotators

---

Delete all code you see there and replace it with:
{
    "query": "How to cook pancakes"
}

Note: you need to delete two brackets {}

Note: you can also find new code by the link

4.3. Leave other interface settings at default.

Data specification ?

Input data	Output data	< >
query (string) ●	**domain** (string) ●	
	intent (string) ○	

⊙ **Define data specification manually**
If the generated specification doesn't work for your task, define custom input and output data. Automatic update of specs due to task interface changes will be disabled. **Learn more**

Show common interface elements

5. Write your own short and simple instructions.

Or you can take already prepared instruction.

Note: instruction by link is in HTML format, so you should enter **source** mode
5.1. Enter **source mode**

Instructions for Tolokers
Write clear and simple instructions

Tolokers prefer to choose tasks that have clear instructions. Describe each step and include examples. The quality of your instructions will directly affect the speed and quality of data labeling. To see examples of good instructions, go to the Knowledge Base
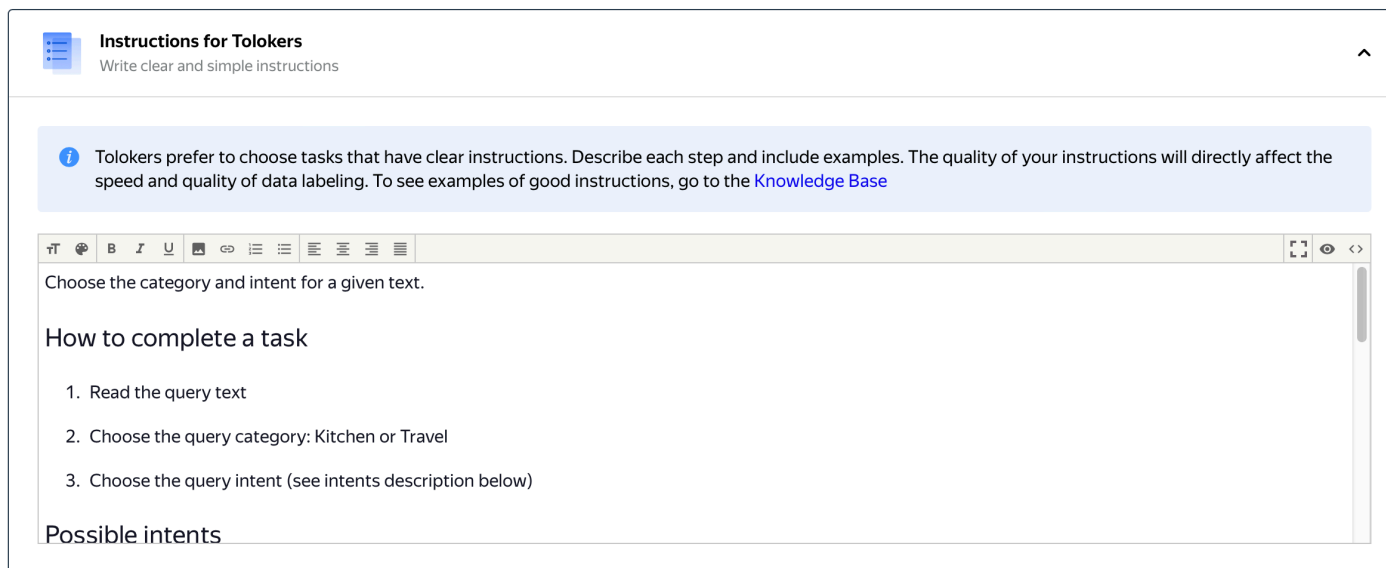
5.2. Paste HTML instruction from link above

5.3. Exit **source mode**

Instructions for Tolokers
Write clear and simple instructions

Tolokers prefer to choose tasks that have clear instructions. Describe each step and include examples. The quality of your instructions will directly affect the speed and quality of data labeling. To see examples of good instructions, go to the Knowledge Base

```
for the <b>travel</b> category, the possible intents are:
<ul>
<li><b>Book a hotel</b> - everything related to booking a flight or hotel</li>
<li><b>Timezone</b> - intention to to get time zone information</li>
<li><b>Travel suggestion</b> - query related to a travel recommendation</li>
<li><b>Exchange rate</b> - intention to find out the price for currency exchange</li>
<li><b>Vaccine</b> - a query containing information about vaccines in a specific region</li>
</ul>
<h2>Example</h2>
<b>Text</b>: "are bananas rich in calories"<br>
<b>Category</b>: Kitchen<br>
<b>Intent</b>: Nutritional information<br>
```

Now you can see formatted instruction

Instructions for Tolokers
Write clear and simple instructions

Tolokers prefer to choose tasks that have clear instructions. Describe each step and include examples. The quality of your instructions will directly affect the speed and quality of data labeling. To see examples of good instructions, go to the Knowledge Base

Choose the category and intent for a given text.

How to complete a task

1. Read the query text

2. Choose the query category: Kitchen or Travel

3. Choose the query intent (see intents description below)

Possible intents

6. Click **Save** to save the project

Edit project

‹› How to work via API   Cancel   **Save**

7.**Note.** To edit project parameters, click the button in the list of projects

Search projects

Create a project

Title ↕	Status	Private comment	Created ↓	Processing
📌 Text classification (category, intent) Self-service	Active		29 Mar, 2023	

Edit

Create a training pool

Text classification (category, intent)
ID 137892

Active

Pools | Training | Statistics | Quality control

1. Click on the **Training** tab and then click **Add training**.

Add training

Use training pools to improve quality. Only Tolokers who complete your training will have access to your paid tasks

Add training

Note: Training is an essential part of almost every crowdsourcing project. It allows you to select performers who have really mastered the task, and thus improve quality. Training is also a great tool for scaling your task because you can run it any time you need new performers.

Training

Instructions

☑ Use project instructions ❓

2. Use the existing project instructions.

Choose the category and intent for a given text.
How to complete a task

1. Read the query text

2. Choose the query category: Kitchen or Travel

3. Choose the query intent (see intents description below)

Possible intents
For **kitchen** category, the possible intents are:

- **Restaurant review** - everything related to whether it is worth visiting a restaurant

- **Restaurant reservation** - intention to book a table in a restaurant

- **Nutritional information** - intention to find information about how healthy the food is or about the number of calories

- **Recipe** - everything related to how to cook food or what it consists of

General settings

Training title	Text classification training ✕
Price per task	☑ free training
Adult content	No
Time per task suite	7200 ✕ seconds
Retry after	10 ✕ days

Task assignment settings

☑ Assign in order of uploading ⍰

☑ Shuffle on page ⍰

Settings for passing training ⍰

Full completion — **Yes**

Required to pass — Number of pages

Create training

3. Specify the training pool settings:

3.1 And then click **Create training**.

---

4. Once the pool is created you can upload tasks. Press Upload. You can upload the training tasks straight from the file from our example which is here.

Note: It's important to include examples for all classes in the training. When running your own projects make sure the training set is balanced and the comments explain why an answer is correct. Don't just name the correct answers.

▶ Text classification training — waiting for data

Statistics | ⬇ Download results | Edit ▾

Download the sample file, add your task data, and upload the file to the pool.
The sample file uses TSV format, which is the same as CSV but with a tab as the separator.
Make sure you choose UTF-8 encoding when saving the file. Learn more

📄 Template for general tasks.tsv

📄 Template for control tasks.tsv

📄 Template for training tasks.tsv

⬆ Upload

0 task pages	0 training tasks
0 tasks	0 control tasks

0 Tolokers who completed training

File upload settings

Tasks per suite

By empty row   Set manually   **Smart mixing**

Training tasks   6

Show advanced settings

Sample file for uploading tasks   Close   Upload

Upload tasks to pools automatically through the API

More details

5. In the opened window choose **Smart mixing**. It is possible to make 1 or several task pages. It seems just reasonable to make 6 task pages.

Adding tasks to pool (training_tasks.tsv)

TASKS FOR POOL

12 tasks   **12** training tasks

0 control tasks

Cancel   Add

6. After the file has been uploaded press **Add** to complete the pool creation.

7. Start the training

Projects › Text classification (category, intent) › Text classification training

⏸ **Text classification training** — paused

Statistics | ⬇ Download results | Edit ⌄

Download the sample file, add your task data, and upload the file to the pool.
The sample file uses TSV format, which is the same as CSV but with a tab as the separator.
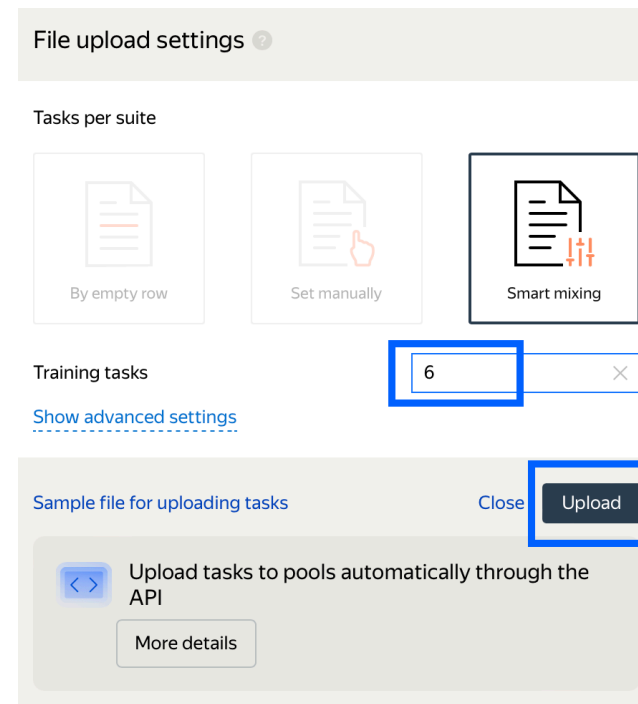Make sure you choose UTF-8 encoding when saving the file. Learn more

📄 Template for general tasks.tsv

📄 Template for control tasks.tsv

📄 Template for training tasks.tsv

⬆ Upload | 📄 Files   👁 Preview

∞ task
   pages

12 training
   tasks

0 tasks

0 control
   tasks

0 Tolokers who completed
   training

Toloka

Pool creation

1. Now go back to our project and click on **Pools** and then **add a pool**

Note: to get back to the project page use these transitions Projects > Project_name > Pool_name

2. Choose the pool type - General tasks. Give the pool any convenient name. You are the only one who will see it.

Text classification (category, intent)
ID 137892

Active

Edit

Key statistics for 7 days ⌄

Submitted tasks	Spent	Control tasks quality	Training tasks quality	Avg. submit time	Tolokers	Banned Tolokers
0	$ 0	No data for the period	No data for the period	No data for the period	0	0

Pools Training Statistics Quality control

Active and closed Archived   Filters   Search   Add a pool

Pools can be archived manually or automatically (automatic archiving applies to pools with no activity for 30 days)

Title ⇕	Priority ⇕	Progress	Status ⇕	Started ⇕	To be completed

To launch a project, you first need to add a pool, set user filters and quality control rules, and upload tasks

50 ⌄

New pool

×

Pool type ⍰

General tasks   ⌄

Pool name (private) *

Text classification (category, intent)   ⊗

Pool description (public) ⍰   ☑ Use project description

Look at the text query and choose the category and intent for it.

╋ Private comment ⍰

Cancel   Create

Get your pool ready for labeling

👤 Select the audience for your tasks ⟩

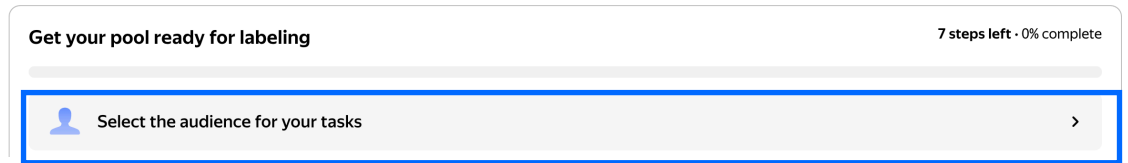# Define relevant tolokers who get access to your tasks

3. Now we need to filter an **audience.**

Filter performers who can access the task. Make sure that 'My tasks may contain shocking ...' section looks like on the screenshot

Choose the Languages options in the list.

You can also set any filters you prefer

Specify the percentage of top-rated performers in the Speed / quality balance.

ℹ️ Filter your audience by language or country. Otherwise, Tolokers anywhere in the world can access your tasks. You can copy audience filters and quality control settings from another pool. Learn more

☐ My tasks may contain shocking or pornographic content. Learn more

| Languages ⌄ | | = | English | ⊗ | 🗑 | + |

☑ Tolokers who passed the language test ❓

[+ Add filter] [+ Add skill]

Speed/quality balance
Note that fewer users means slower pool completion
Learn more

[Top %] [Online]

Specify the percentage of top-rated Tolokers who can access tasks in the pool

8144 🧍 ─────●──────────────────────────────── 814 🧍
Speed   100% 90%   80%   70%   60%   50%   40%   30%   20%   10% Quality

90% of Tolokers selected
7329 Tolokers available

[Save] [Cancel]

✳️ Toloka

4. Set up **Quality control**.

Attach the training you created earlier and select the accuracy level that is required to reach the main pool.

Note: This means that Tolokers who get less than 85% accuracy will not see this pool.

5. Update the **Fast responses** rule

Note: This rule allows you to ban performers who submit tasks at a suspiciously high speed.

Get your pool ready for labeling
6 steps left · 14% complete

✓ Select the audience for your tasks ›

🛡 Set up quality control ›

Quality rules to protect project

To get more accurate responses, select quality control rules. All rules work independently of one another

⊘ Review task responses manually ❓

Training	Level required, % *
Text classification training ⌄	85

FAST RESPONSES ❓

Recent task suites to use	5 ✕
Minimum time per task suite	10 ✕

If number of fast responses ⌄ ≥ 1 ✕ +
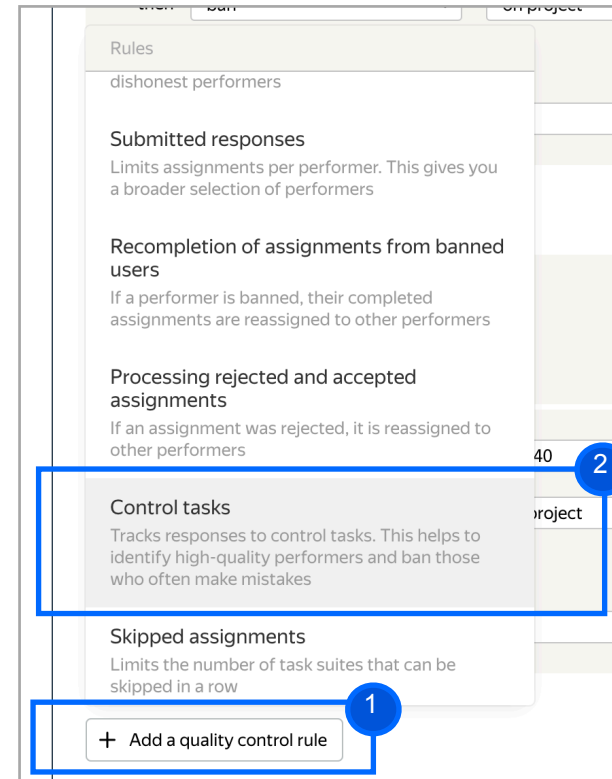
then ban ⌄

on project ⌄

1 ✕ day ⌄

Fast responses ✕

Save Cancel

6. Let's add one more role: **control tasks**

Click **Add a quality control rule** and choose **Control tasks**.

7. Set the number of responses and the percentage of correct responses. Ban performers who give incorrect responses to control tasks.

Note: Since tasks such as these have an answer that can be used as ground truth, we can use standard quality control rules like golden sets.

Recent control and training task responses to use | items

If correct control responses (... ⌄ < 80 ✕ +

and number of control respons... ⌄ > 3 🗑 +

then ban ⌄

on project ⌄

3 ✕ days ⌄

Control tasks ✕

➕ Add a quality control rule

Save Cancel

8. Set the **price** per task suite (for example, $0.01).

Also, choose an **overlap** of 3.

Click **Save**

Price and overlap

📖 How to estimate a fair price

Price per task suite, $ 0.01

▮▮ **Doesn't meet Toloker expectations**
At this price, Tolokers will earn less than minimum wage in the countries you selected. To improve interest in your tasks, increase the price to 0.292 per task suite

Increase price

Number of Tolokers per task Overlap 3
To get reliable quality from aggregated responses, we recommend 3 Tolokers per task. Learn more

Flexible price and overlap ⌄

✳ Toloka

Save Cancel Your price per 1 item is **$0.004** ⍰

9. We need to upload our main data. Please take it here and load it

Click **Continue**

Get your pool ready for labeling

4 steps left · 43% complete

✓ Select the audience for your tasks >

✓ Set up quality control >

✓ Set the task price and overlap >

⚙ Add optional pool settings   Leave default >

☁ Prepare and upload data   Skip or upload via API >

Upload data

Download the template for general tasks and add your data

XLSX
Template in XLSX

TSV
Template in TSV

JSON
Template in JSON

or

⬇ Templates for control tasks and training

<> Learn how to do it via API

Attach the prepared file with data

main_and_golden_train_tasks.tsv   ✓ 468 items ✕

Continue   Cancel

10. You need to determine how many tasks do you want per suite

Let's say we want **5 general tasks and 1 control task**s

We recommend putting as many tasks on one page as an annotator can complete in 1 to 5 minutes. That way, annotators are less likely to get tired, and they won't lose a significant amount of data if a technical issue occurs.

Click **Combine tasks into suites**

How many tasks do you want per suite?

Smart mixing	Set manually

Set how many tasks of each type to mix in each task suite. For better quality, include control tasks. You can mark up control tasks on the pool page after uploading data, or upload them in a separate file and adjust the smart mixing settings later. Learn more

General tasks *	Control tasks *	Training tasks *
5	1	0

Show advanced settings ⌄

Combine tasks into suites   Back

That's it! Let's do final steps
and go on with project!

Get your pool ready for labeling

2 steps left · 71% complete

✓ Select the audience for your tasks ›

✓ Set up quality control ›

✓ Set the task price and overlap ›

⚙ Add optional pool settings · Leave default ›

✓ Prepare and upload data ›

✓ Add control tasks for checking performance ›

☑ Double-check your project and try out tasks · Do it later ›

Toloka

11. Click on **Start labeling**

You can finally check estimated expenses and launch labeling

Setup is finished and your pool is ready for labeling

▶ Start labeling

Toloka | **Projects** | Users | Skills | Profile | Messages 2208 | Docs & Resources

Projects › Text classification (category, intent) › Text classification (category, intent)

General tasks

Text classification (category, intent) ✎

Public description: Look at the text and choose the category and intent for it.

Setup is finished and your pool is ready for labeling

Overview | Efficiency | Statistics | Settings

⚠ Some issues need your attention
Check our recommendations to improve efficiency

Tasks ＋ ⋯

~270
Task suites

450
Tasks

Audience 👤

Speed/quality balance 90% top-rated Tol

Languages English (exam-ba

Review budget and launch

Text classification (category, intent)

Pricing

Price per task suite	$0.01
Task price	$0.002
Overlap	3 Tolokers

Task details

Task suites	270
Tasks	468

Subtotal	$2.70
Fee, 20%	$0.54
Total	**$3.24**

▶ Launch

※ Toloka

We have just launched the labeling of the objects on which we will **train** our classification model.

We also need to label the **test** dataset to validate the resulting model.

12. Return to the list of pools

13. Clone your pool

14. Set a different name for your convenience

General tasks

Text classification (category, intent) ✏️

Public description: Look at the text and choose the category and intent for it.

⏳ **Tolokers are completing your tasks**
Start date: March 29, 2023 5:55 PM

⬇️ Downl

Pools   Training   Statistics   Quality control

Active and closed   Archived   Filters   [Search]   Add a pool

Pools can be archived manually or automatically (automatic archiving applies to pools with no activity for 30 days)

Title ⇅	Priority ⇅	Progress	Status ⇅	Started ⇅	To be completed
Text classification (category, intent)	0	▭ 0 / 270	Labeling...	March 29, 2023 5:55 PM	⏸️ ✏️ ⧉

Clone pool   ✕

Pool type ❓

General tasks   ⌄

Pool name (private) *

Test dataset   ⊗

Pool description (public) ❓   ☑️ Use project description

Look at the text and choose the category and intent for it.

＋ Private comment ❓

Cancel   Clone

✳️ Toloka

15. Download data for labeling from this link and upload it to the newly cloned pool

Get your pool ready for labeling

3 steps left · 57% complete

✓	Select the audience for your tasks	›
✓	Set up quality control	›
✓	Set the task price and overlap	›
✓	Add optional pool settings	›
☁	Prepare and upload data	Skip or upload via API ›

Upload data

Download the template for general tasks and add your data

XLSX	TSV	JSON
Template in XLSX	Template in TSV	Template in JSON

or

↓ Templates for control tasks and training

‹› Learn how to do it via API

Attach the prepared file with data

📁 Drop file here or select
Format: XLSX, TSV, JSON

16. Since we cloned pool, all settings are the same, so we don't need to change anything

How many tasks do you want per suite?

Smart mixing | Set manually

Set how many tasks of each type to mix in each task suite. For better quality, include control tasks. You can mark up control tasks on the pool page after uploading data, or upload them in a separate file and adjust the smart mixing settings later. Learn more

General tasks *
5

Control tasks *
1

Training tasks *
0

Show advanced settings ⌄

17. Check your project or skip it and start labeling

Combine tasks into suites | Back

General tasks

Test dataset ✎

Try tasks as a Toloker | ✎ | ⋯

Public description: Look at the text and choose the category and intent for it.

✓ Setup is finished and your pool is ready for labeling

▶ Start labeling

✳ Toloka

1. Wait until the pool is completed.

Pools can be archived manually or automatically (automatic archiving applies to pools with no activity for 30 days)

Title ⇕	Priority ⇕	Progress	Status ⇕	Started ⇕	To be completed
Test dataset	0	1 / 162	Labeling...	March 29, 2023 6:13 PM	~ 2 hours
Text classification (category, intent)	0	270 / 270	Labeled	March 29, 2023 5:55 PM	—

2. Click **Download results**

Labeling completed!
Start date: March 24, 2023 2:55 PM

View assignments   ⬇ Download results ▾

Accepted **100%** (270 of 270)

51	51	0	3	$2.70 (+ $0.54)
Interested Tolokers	Tolokers submitted task	Skipped task suites	Expired task suites	Labeling costs (+ fees)

3. You can choose which tasks you want to upload. But leave the **default settings**

Click **Download**

✕

Download results

[All results] [For period]

Task type

☑ General   ☑ Control   ☑ Training

Task status

☐ Active   ☐ Submitted   ☑ Accepted

☐ Rejected   ☐ Skipped   ☐ Expired

Additional options ⌄

Cancel   **Download**

✳ Toloka

25

You can find all files from this tutorial by the link:
https://drive.google.com/drive/folders/1jMNkCs1DzJESiL2-8Pr8NJhqyi7GZJp5?usp=sharing