



Intent classification manual

Create a project

1. Click **Create a project**.

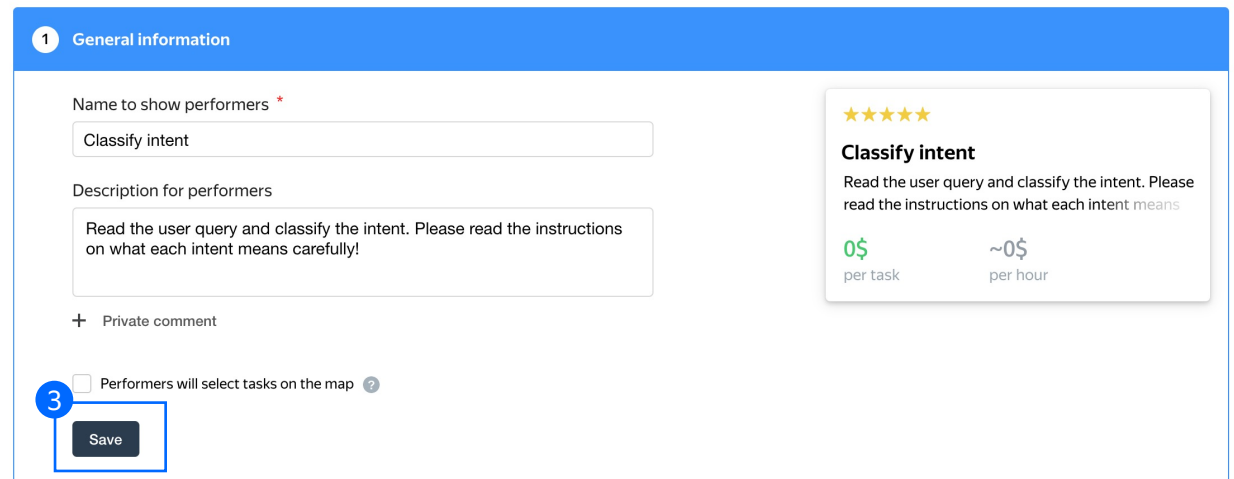
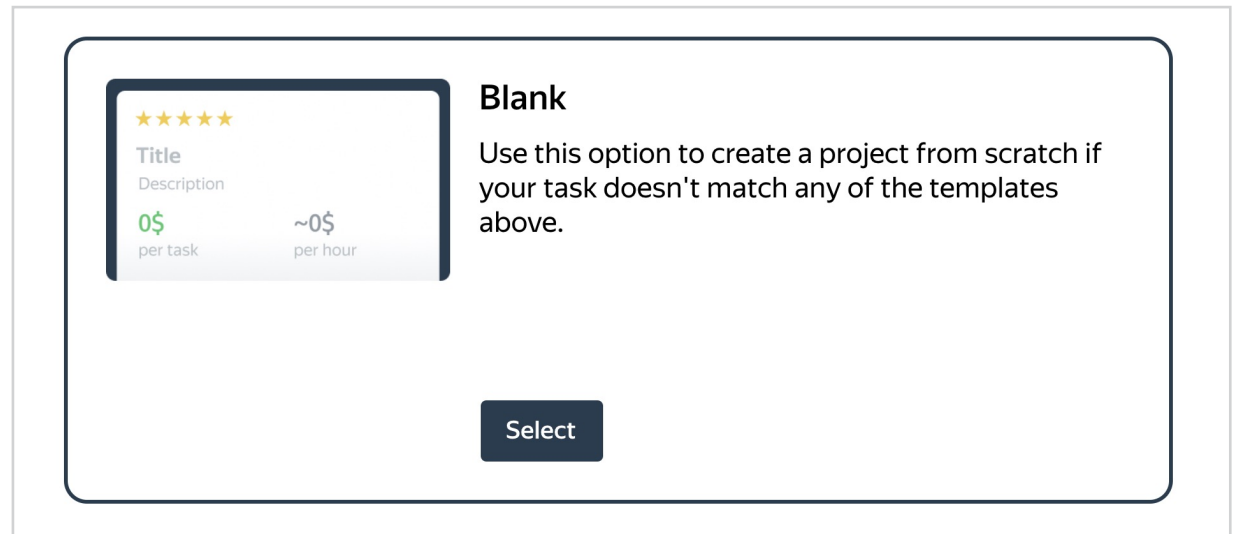
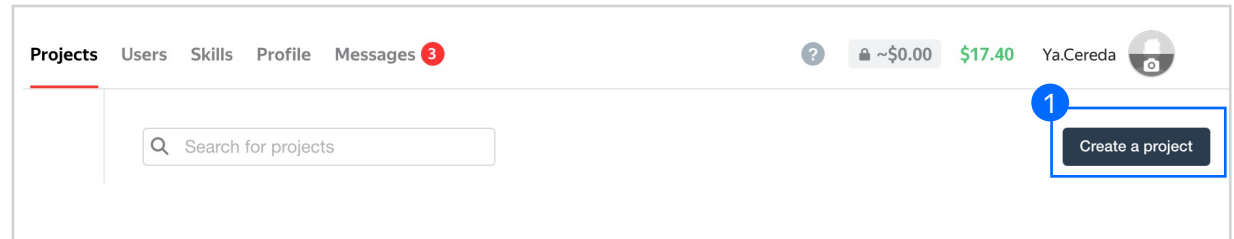
2. Choose the **Blank** template.

You can also use the **Text Classification** preset and update it according to your task specifics.

Read more about configuring the [task interface](#) in the Requester's Guide.

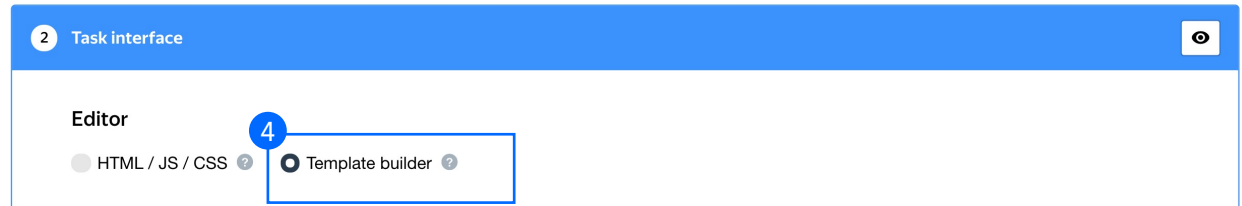
3. Enter a clear project name and description. Click **Save**.

Note: The project name and description will be visible to the performers.



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

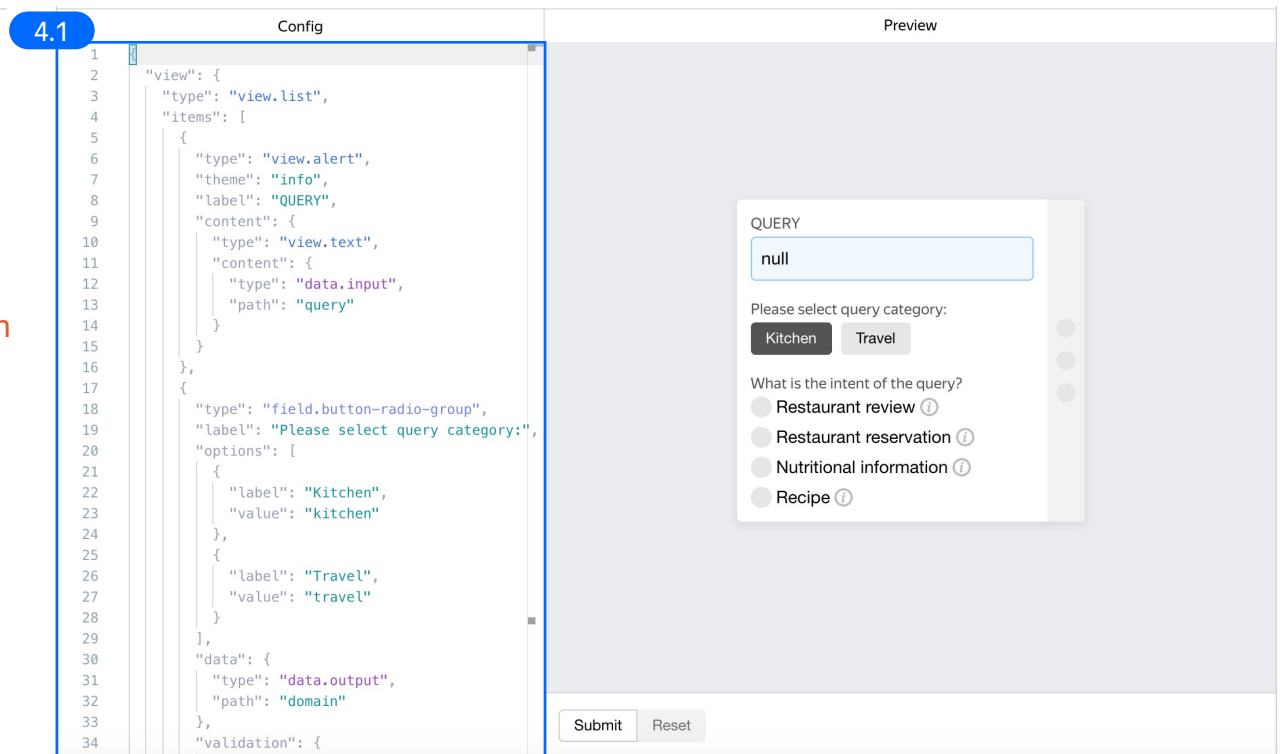
Read more about the [Template builder](#) in the Requester's Guide.



- 4.1. Delete the existing config and paste the code provided at the end of this manual (in the appendix).

Note: You can check what the template looks like and how it works in the preview section. For this task, we added short reminders about each intent. They appear if a user clicks on the **Info** sign. This is a simple yet helpful feature that helps performers choose answers.

Check the [Interfaces section](#) of our Knowledge Base for more tips on interface design.



4.2. Make sure the specifications look like this:

Note: Specifications are a description of input data that will be used in a project and the output data that will be collected from the performers.

Read more about [input and output data specifications](#) in the Requester's Guide.

The screenshot shows the 'Data specification' interface. It has two main panels: 'Input data' and 'Output data'. The 'Input data' panel contains a field labeled 'query (json)' with a radio button. The 'Output data' panel contains two fields: 'domain (string)' with a radio button and 'intent (string)' with a radio button. Below each panel is an 'Add field' button. At the bottom of the interface, there is a 'Show common interface elements' button and a 'Save' button. A blue circle with the number '4.2' is placed over the 'Save' button.

5. Write comprehensive instructions. Be sure to add descriptions for every class and subclass, as well as examples.

Click **Save**.

Get more tips on designing [instructions](#) in our Knowledge Base.

The screenshot shows the 'Instructions for performers' interface. It has a blue header with the title 'Instructions for performers'. Below the header, there is a text area with instructions: 'When a performer selects a task, they first see the instructions that you wrote. Describe what needs to be done and give examples. You can prepare your instructions in HTML format, then copy and paste them into the editor. Press < > to switch to HTML mode. To learn more, see the [documentation](#).' Below the text area is a rich text editor with a toolbar. The editor contains the following text: 'Please , choose the right category and related class for given intent. Category : 1. Kitchen 2. Travel Kitchen class description 1. Restaurant review - type of intent where it is asked about restaurant estimation: food quality, number of reviews, review to the restaurant.' At the bottom of the editor is a 'Save' button. A blue circle with the number '5' is placed over the 'Save' button.

6. Leave the **Translations** block as default and click **Save**.

4 Translations

Performers from different countries will understand the purpose of the task better if it's in their own language. Translate the task name, description, and instructions into each language that you want performers to see. Otherwise, the language is inactive. If you want to translate the task interface, you have to create it using Template Builder.

Source language
—

Translations

Language	Name and description for performers	Instructions for performers	Task interface
✓ Source	✓	✓	✓

Add translation

6 Save

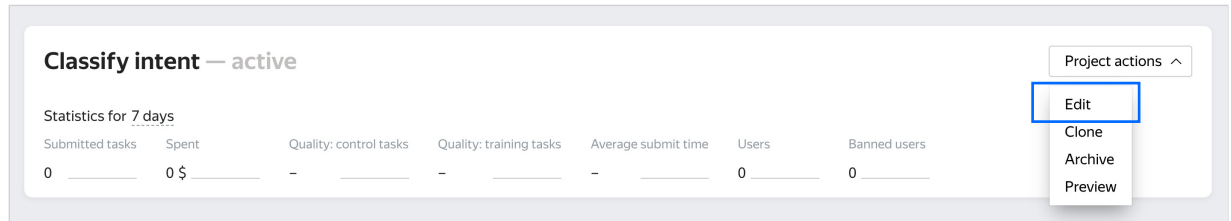
7. Click **Finish** to save the project.

New project

[Back to the old interface](#) Cancel **7** Finish

- ✓ General information
- ✓ Task interface
- ✓ Instructions for performers
- ✓ Translations

Note: To edit project parameters, click the button in the list of projects or go to **Project actions** → **Edit** on the project page.



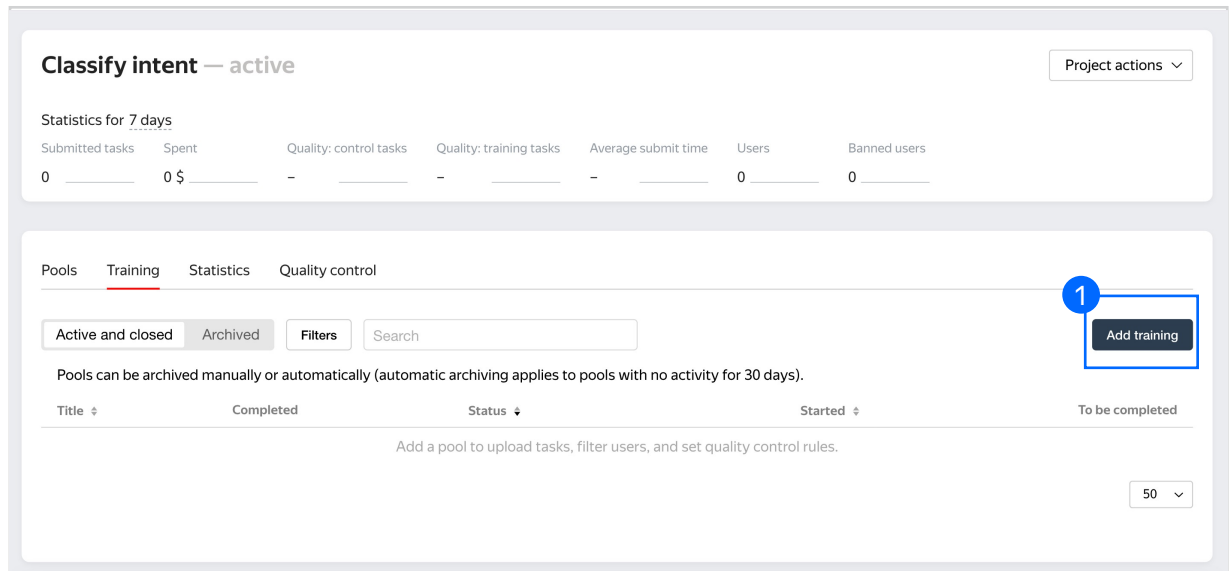
The screenshot shows the 'Classify intent' project page, which is active. At the top right, there is a 'Project actions' dropdown menu. The 'Edit' option is highlighted with a blue box. Below the menu, there is a statistics section for the last 7 days, showing metrics like Submitted tasks, Spent, Quality: control tasks, Quality: training tasks, Average submit time, Users, and Banned users. The 'Users' metric is currently at 0.

Create a training pool

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

Read more about [selecting performers](#) in our Knowledge Base.



The screenshot shows the 'Classify intent' project page, which is active. The 'Training' tab is selected. In the top right corner, there is a 'Project actions' dropdown menu. Below the menu, there is a statistics section for the last 7 days, showing metrics like Submitted tasks, Spent, Quality: control tasks, Quality: training tasks, Average submit time, Users, and Banned users. The 'Users' metric is currently at 0. In the bottom right corner, there is a blue box with the number '1' and the text 'Add training'.

2. Use the existing project instructions.

Instructions

☒ Use project instructions ?

Please , choose the right category and related class for given intent.

Category :

- 1. Kitchen
- 2. Travel

3. Specify the training pool settings:

3.1. Click **Create training**.

Read more about [training pools](#) in our Requester's Guide.

General settings

Training title

Price per task ☒ free training

Adult content ☐ No

Time per task suite seconds

Retry after days

Task assignment settings

☒ Assign in order of uploading ?

☒ Shuffle on page ?

Settings for passing training ?

Full completion ☒ Yes

Required to pass


3.1


4. Upload training tasks to the pool.
You can download a sample file with tasks [here](#).


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

File upload settings ?

Tasks per page


By empty row


Set manually


Smart mixing

Training tasks

[Show advanced settings](#)

Sample file for uploading tasks

Close

4 Upload

If the upload is successful, you will see the updated task counter.

Upload

Files

Delete

Edit

Preview

∞

task pages

12

training tasks

0

tasks

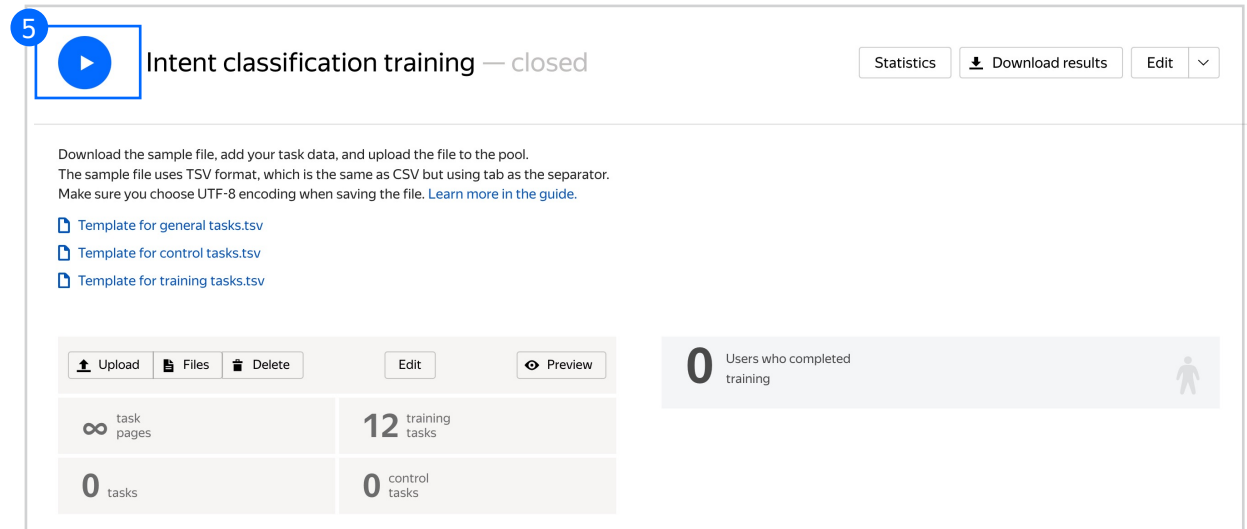
0

control tasks

5. Run the training pool.

Note: We recommend opening the training pool along with the main pool. Otherwise Tolokers will spend their time on training but get no access to real tasks, which is frustrating.

Also, do not forget to close the training pools when there are no main tasks available anymore.

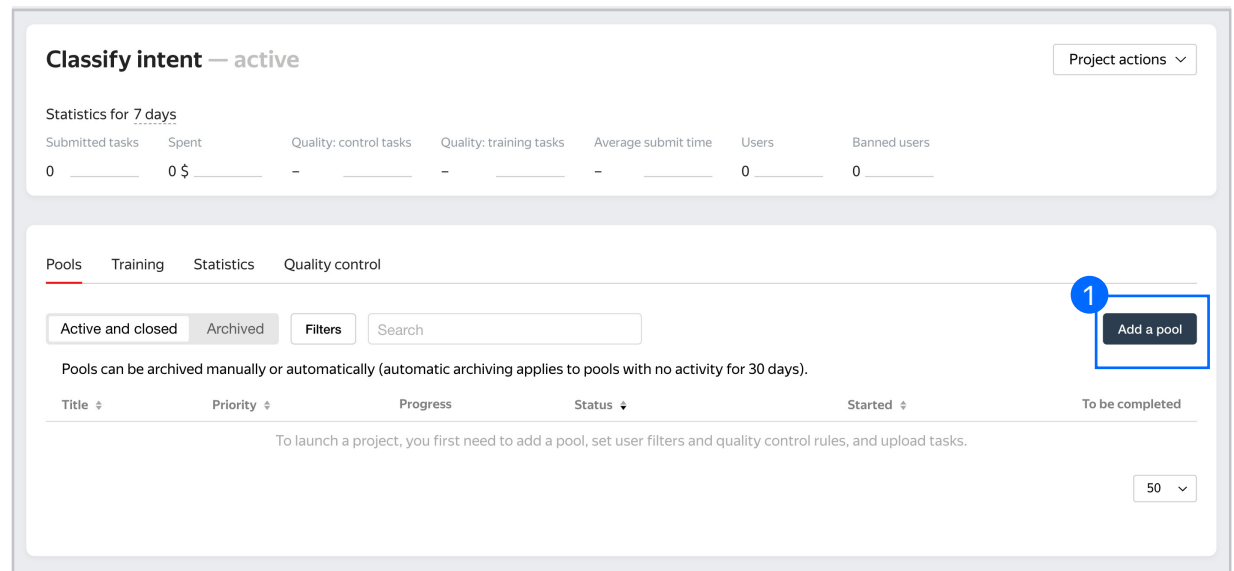


Create the main pool

1. Click Add a pool.

A pool is a set of paid tasks grouped into task pages. These tasks are sent out for completion at the same time.

Note: All tasks within a pool have the same settings (price, quality control, etc.)



2. Give the pool any name you find suitable. You are the only one who will see it.

The description can be either public or private. Choose the option you prefer.

2

POOL NAME (VISIBLE ONLY TO YOU) ?

☒ Use project description

PUBLIC DESCRIPTION ?

Read the user query and classify the intent. Please read the instructions on what each intent means carefully!

☐ Add a private description

3. Specify [pool parameters](#):

- 3.1. Select pool type. Choose **Other**.

Read more about [pool types](#) in the Requester's Guide.

POOL TYPE ?

Other ^

—

Exam

Training

Retry

✓ Other

PRICE IN US DOLLARS ?

FEE ? 0.005

- 3.2. 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.

3.2

TRAINING

LEVEL REQUIRED, %

3.3. Set the price per task suite
(for example, \$0.01).

Classification tasks are normally paid as basic tasks because these tasks do not take much time.

Read more about [pricing principles](#) in our Knowledge Base.

3.3

PRICE IN US DOLLARS ?

0.01



FEE ? 0.005

+ Dynamic pricing

Price per task suite

Each task suite can have one or multiple tasks on the same page. Enter the total price for all tasks in the suite.

3.4. [Filter](#) performers who can access the task. Choose “No” in the **Adult content** block.

Click **Add filter** to choose the **Languages** and **Client** options in the list.

3.4

ADULT CONTENT ?

No

Add filter

[Copy settings from...](#)

Performers

Filter performers who can access the task.
Toloka has users from different countries, so don't forget to filter by language and region. [Learn more](#)

Add skill

3.5. Choose **Languages = English** as your first filter. This way, performers who speak English will be invited to complete this task.

Then choose **Toloka web version** and **Toloka for mobile** clients.

These filters will make it possible for performers to complete your task on their computers or mobile devices.

Add filter ▾ Add skill

3.5 PERFORMER PROFILE

Languages ▾ = English ✕

☐ Performers who passed the language test ?

AND

CALCULATED DATA

Client ▾ = Toloka web version ▾

OR = Toloka for mobile ▾

Calculated data

- Region by phone number
- Region by IP
- Performer rating
- Client
- Device type
- Operating system
- OS versions
- OS major version
- OS minor version

3.6. Set up [Quality control](#). Ban performers who give incorrect responses to control tasks. Click + **Add Quality Control Rule**.

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.

Read more about [quality control principles](#) in our Knowledge Base or check out [control task settings](#) in the Requester's Guide.

Quality control

Add rules to get more accurate responses.
All rules work independently.

NON-AUTOMATIC ACCEPTANCE ? ☐ No

REVIEW PERIOD IN DAYS 14

CAPTCHA FREQUENCY ? None ▾

3.6 + Add Quality Control Rule

3.7. Click **Control tasks**. Set the number of responses and the percentage of correct responses.

Recent control task responses to use

3.7

If > +

and < +

then on requester days

3.8. Set up the up the [Fast responses rule](#).

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

Recent task suites to use

3.8

Minimum time per task suite

If > +

then on requester forever

3.9. Overlap. This is the number of users who will complete the same task.

Set an overlap of 3 to get a more confident final label.

To understand [how this rule works](#), go to the Requester's Guide.

Overlap

Specify how many performers you want to complete each task in the pool.

3.9

OVERLAP ? 3

DYNAMIC OVERLAP ? ☐ Off

3.10. Optionally, specify the percentage of top-rated performers in the [Speed / quality ratio](#).

Note: This can slow down pool completion.

Speed/quality balance

Set additional filters to restrict performer access based on their rating in Toloka. This boosts quality but may slow down project completion because there will be fewer performers available to complete tasks. [Learn more...](#)

Top %

Online

Time

Specify the percentage of top-rated active users who can access tasks in the pool.

3.10

8235

Speed

All

90%

80%

70%

60%

50%

40%

30%

20%

10%

5764

Quality

70% top-rated performers were selected.
The task is available to **5764** active users.

3.11. Specify the time given to complete a task suite (for example, 600 seconds).

To understand how much time it should take to complete a task suite, try doing it yourself.

3.11

Parameters

TIME PER TASK SUITE IN SECONDS ? 600 X

POOL CLOSING DATE ? 2022-08-04

KEEP TASK ORDER ? ☐ No

WAITING TIME FOR THE POOL TO CLOSE IN SECONDS ? 10 X

POOL PRIORITY WITHIN THE PROJECT ? 0

4. Click **Save** to save Pool parameters.

Parameters

TIME PER TASK SUITE IN SECONDS ? 600 X

POOL CLOSING DATE ? 2022-08-06

KEEP TASK ORDER ? ☐ No

WAITING TIME FOR THE POOL TO CLOSE IN SECONDS ? 0

POOL PRIORITY WITHIN THE PROJECT ? 0

Cancel

4 Save

Prepare and upload a file with tasks

1. Prepare a TSV file with tasks as shown in our [example](#). You can also upload a [file](#) that already contains golden sets. [Origin](#)

```
@inproceedings{larsen-etal-2019-evaluation,
title = "An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction",
author = "Larsen, Stefan and Mahendran, Anish and Peper, Joseph J. and Clarke, Christopher and Lee, Andrew and Hill, Parker and Kummerfeld, Jonathan K. and Leach, Kevin and Laurenzano, Michael A. and Tang, Lingjia and Mars, Jason",
booktitle = "Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)",
year = "2019",
url = "https://www.aclweb.org/anthology/D19-1131" }
```

License: CC BY 3.0

2. [Upload pool tasks](#) from this file.

- 2.1. Select [Smart mixing](#) in **File upload settings** and specify the number of tasks of each type per page. Click **Upload**.

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

To learn more about [grouping tasks](#) into suites, read the Requester's Guide.

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

Classify intent — closed

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 using tab as the separator. Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

- Template for general tasks.tsv
- Template for control tasks.tsv
- Template for training tasks.tsv

2 Upload

0 task pages	0 training tasks
0 tasks	0 control tasks

0 % Completed 0

File upload settings ?

Tasks per page

By empty row Set manually Smart mixing

Main tasks 3

Training tasks 0

Control tasks 1

[Show advanced settings](#)

Sample file for uploading tasks Close Upload

2.1

3. [Create control tasks.](#)

Click **Edit** → **Create control tasks**.

Note: you can prepare control tasks not only in the interface. You can collect them outside Toloka and upload a task set with golden answers: [example](#).

Classify intent — closed

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 using tab as the separator. Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

[Template for general tasks.tsv](#)
[Template for control tasks.tsv](#)
[Template for training tasks.tsv](#)

Upload Files Delete Edit

0 task pages 0 training tasks
162 tasks 0 control tasks

0 % Completed 0

Note: Control tasks are tasks that already contain the correct response. They are used for checking the quality of responses from performers. The performer's response is compared to the response you provided. If they match, it means the performer answered correctly.

Projects > Classify intent > Classify intent > Uploaded tasks

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 160 Control tasks 0 Training tasks 0

Create control tasks Create training tasks Download

ID	Overlap	Responses from performers	Last updated
...079acc1e	3	0	08/05/2021 4:01:56 PM
...079acc27	3	0	08/05/2021 4:01:56 PM
...079acc26	3	0	08/05/2021 4:01:56 PM
...079acc23	3	0	08/05/2021 4:01:56 PM
...079acc28	3	0	08/05/2021 4:01:56 PM
...079acc22	3	0	08/05/2021 4:01:56 PM

3.1. Enter correct responses to your control tasks. Check the result output field, which compares the user's response to the control task. Look at the query, select the response, click **Save** and go to **next**. Repeat until you have ~20 control tasks.

In small pools, control tasks should account for 10–20% of all tasks.

To learn more about [creating control tasks](#), go to the Requester's Guide.

Projects > Classify intent > Classify intent > Uploaded tasks > Edit tasks

Create 2 more

Create control task

1. Enter correct responses

2. Select the fields to use

Field	Value
<input checked="" type="checkbox"/> domain	
<input checked="" type="checkbox"/> intent	

QUERY

Is chocolate bad for you or good for you

Please select query category:

Kitchen Travel

Save and go to next

3.1

Distribution of correct responses for control tasks

Create control tasks to see charts of response distribution.

Tip. Make sure to include different variations of correct responses in equal amounts.

Go to **Uploaded tasks** → **Control tasks** → **Distribution of correct responses for control tasks** tab.

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 162 Control tasks 18 Training tasks 0

Create from main tasks Download

ID	Overlap	Responses from performers	Correct responses, %	Last update
...f0920730	∞	14	100	08/04/2021 11:58
...f092070b	∞	11	100	08/04/2021 11:58
...f092071a	∞	9	77.778	08/04/2021 11:58
...f0920723	∞	9	100	08/04/2021 11:58
...f0920709	∞	5	60	08/04/2021 11:58
...f09206fa	∞	8	100	08/04/2021 11:58
...f09206f4	∞	10	90	08/04/2021 11:58
...f09206e4	∞	10	90	08/04/2021 11:58

Distribution of correct responses for control tasks

intent

- 11.1% bookhotel
- 11.1% exchange_rate
- 11.1% nutrition
- 11.1% recipe
- 11.1% reservation

Show other values

domain

- 47.1% kitchen
- 52.9% travel

3.2. To check the number of control tasks, go to the **Pool** page.

Classify intent — closed

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 using tab as the separator. Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

[Template for general tasks.tsv](#)
[Template for control tasks.tsv](#)
[Template for training tasks.tsv](#)

Upload Files Delete Edit Preview

~162 task pages	0 training tasks
162 tasks	18 control tasks

0 % Completed 0

0 ~162

4. Preview the pool.

Note: Remember that the tasks will be completed by actual Tolokers. Double check that everything is correct with your project configuration.

Preview pool tasks Back

Tasks Active Messages 18:35 / \$0.01 Classify intent

QUERY: does boston restaurant have favorable reviews
Please select query category: Kitchen Travel

QUERY: how is the steamed fish at harbor village
Please select query category: Kitchen Travel

QUERY: how is dim sum at legend
Please select query category: Kitchen Travel


QUERY: what are the ratings for yardhouse
Please select query category: Kitchen Travel

Exit Skip Submit

5. Start the pool.

Make sure you start the training pool before you start the main pool.

5



Classify intent — closed

StatisticsDownload resultsEdit

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 using tab as the separator.
Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

[Template for general tasks.tsv](#)
[Template for control tasks.tsv](#)
[Template for training tasks.tsv](#)

UploadFilesDeleteEditPreview

~162 task pages	0 training tasks
162 tasks	18 control tasks

0 %
Completed 0

0~162

Receiving responses

1. Wait until the pool is completed.
Refresh the pool page to check progress.

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

2. Click **Download results** and run aggregation using the [Dawid-Skene model](#).

Note: We use this aggregation model because our questions are of comparable difficulty, and we don't have many control tasks.

Read more about the [Dawid-Skene model](#) in the Requester's Guide or get at an overview of different [aggregation models](#) our Knowledge Base.

3. Go to the list of operations and wait until aggregation is complete.
Refresh the **Operations** page to check progress.

The screenshot shows the 'Classify intent' pool page, which is marked as 'closed'. It features a 'Download results' button and a 'View operations' button. A dropdown menu for 'View operations' is open, showing 'Dawid-Skene aggregation model' and 'Aggregation by skill'. The page displays a progress bar at 100% with the text 'Completed 162, accepted 162'. Below the progress bar is a 'View assignments' button. The page also includes a table with task statistics:

Task Type	Count
task pages	162
training tasks	0
tasks	162
control tasks	18

The screenshot shows the 'Operations' page with a table of operations. The table has columns for Id, Type, Started, Completion time, Progress, Status, and Files. A dropdown menu for 'Progress' is open, showing '0%'. The table contains one row of data:

Id	Type	Started	Completion time	Progress	Status	Files
393e1c...	Dawid-Skene aggregation model	08/06/2021 6:52:24 PM	08/06/2021 6:54:51 PM	0%	Running	-

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

Operations

PROJECT

Classify intent

POOL

Classify intent

Id	Type	Started	Completion time	Progress	Status	Files
393e1c...	Dawid-Skene aggregation model	08/06/2021 6:52:24 PM	08/06/2021 6:54:51 PM	<div></div> 100%	Success	<div>4</div> <div>Download</div>

10

Appendix

Interface code

Step 4.1.

```
{
  "view": {
    "type": "view.list",
    "items": [
      {
        "type": "view.alert",
        "theme": "info",
        "label": "Query",
        "content": {
          "type": "view.text",
          "content": {
            "type": "data.input",
            "path": "query"
          }
        }
      },
      {
        "type": "field.button-radio-group",
        "label": "Please select query category:",
        "options": [
          {
            "label": "Kitchen",
            "value": "kitchen"
          },
          {
            "label": "Travel",
            "value": "travel"
          }
        ],
        "data": {
          "type": "data.output",
          "path": "domain"
        },
        "validation": {
          "type": "condition.required"
        }
      },
      {
        "type": "helper.if",
        "condition": {
          "type": "condition.equals",
          "data": {
            "type": "data.output",
            "path": "domain"
          }
        },
        "for": "kitchen"
      },
      {
        "type": "field.button-radio-group",
        "label": "What is the intent of the query?",
        "options": [
          {
            "label": "Restaurant review",
            "value": "review",
            "hint": "Anything related to whether a restaurant is worth the visit"
          },
          {
            "label": "Restaurant reservation",
            "value": "reservation",
            "hint": "Intention to reserve a table in the restaurant"
          },
          {
            "label": "Nutritional information",
            "value": "nutrition",
            "hint": "Intention to know information about how healthy food or amount of calories "
          },
          {
            "label": "Recipe",
            "value": "recipe",
            "hint": "Anything related to how to prepare food or food substitution"
          }
        ],
        "data": {
          "type": "data.output",
          "path": "intent"
        },
        "validation": {
          "type": "condition.required",
          "hint": "please select the relevant intent"
        }
      },
      {
        "type": "helper.if",
        "condition": {
          "type": "condition.equals",
          "data": {
            "type": "data.output",
            "path": "domain"
          }
        },
        "for": "travel"
      },
      {
        "type": "field.button-radio-group",
        "label": "What is the intent of the query?",
        "options": [
          {
            "label": "Book a hotel",
            "value": "bookhotel",
            "hint": "Anything related to reservation a flight or hotel"
          },
          {
            "label": "Timezone",
            "value": "timezone",
            "hint": "Intention to get information about timezone"
          }
        ],
        "data": {
          "type": "data.output",
          "path": "intent"
        },
        "validation": {
          "type": "condition.required",
          "hint": "please select the relevant intent"
        }
      }
    ]
  },
  "plugins": [
    {
      "type": "plugin.toloka",
      "layout": {
        "type": "scroll",
        "taskWidth": 300
      }
    }
  ]
}
```