

Search relevance evaluation manual

	Create a project	Projects Users Skills Profile Messages 3	a.Cereda
1.	Click Create a project.	Q Search for projects	Create a project
2.	Choose the Blank template. Read about configuring the <u>task</u> <u>interface</u> in the Requester's Guide.	***** Title Description 0\$ ~0\$ per task per hour	
		Select	

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

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

General information	
Name to show performers * Classify search query relevance Description for performers Analyze a website with a product and decide to what extent it meets the search query	***** Classify search query relevance Analyze a website with a product and decide to what extent it meets the search query 0\$ ~0\$
+ Private comment	per task per hour

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

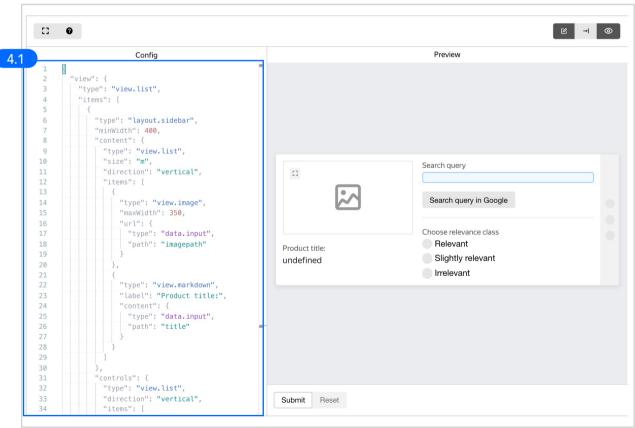
Read more about the <u>Template</u> <u>builder</u> in the Requester's Guide.



4.1. You can either update the existing config or delete it and paste the code provided at the end of this manual (in the appendix).

Note: This interface contains a query, a picture of a product, and its title, which needs to be assessed. There is a button for checking this query in Google, which is handy because the query might not be obvious and performers will often need to look it up. There is also a plugin that checks if a label was really chosen.

Check the <u>Interfaces section</u> of our Knowledge Base for more tips on interface design.

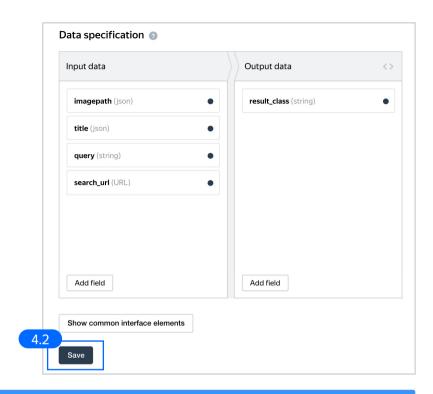


4.2. Make sure the specifications look like this:

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.

Note: We are using screenshots to make this demo more robust against possible webpage changes. Another way is to use an <u>iframe</u> and let the performers assess the whole webpage.

Read more about <u>input and output data</u> <u>specifications</u> in the Requester's Guide.



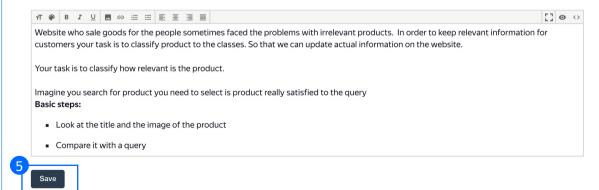
5. Write comprehensive instructions.

Note: Instructions are essential for complex tasks like relevance evaluation that are based on a set of rules and various criteria. Make sure to not only describe the general idea, but also go through examples and explain the evaluation logic in each case. We recommend trying to evaluate around two dozen cases yourself to get more insights for the instructions.

Get more tips on designing instructions in our Knowledge Base.

3 Instructions for performers

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.



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

4	Translations				
	description, ar		you want performers to see. Otherwis	heir own language. Translate the task name, e, the language is inactive. If you want to	
	Source language	~			
	Translations				
	Language	Name and description for performers	Instructions for performers	Task interface	
	Source	\checkmark	\checkmark	\checkmark	
6	Add translation	~			

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

New project	Back to the old interface	Cancel	Finish
General information			
C Task interface			
Instructions for performers			
Translations			

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

Classify search query relevance — active					Project action		
Statistics for 7 d	lavs						Edit
Submitted tasks	Spent	Quality: control tasks	Quality: training tasks	Average submit time	Users	Banned users	Clone
	0\$				0	0	Archive
0	U Ş		-	-	0	U	Preview

Create a training pool

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

Note: Since relevance evaluation is based on rules, not just common sense or certain skills, we recommend investing some time on learning how to explain all the rules. Training needs to involve both common and extreme cases. The comments should explain the underlying logic rather than just state the correct answers.

Note: A well-grounded training exercise is also a great tool for scaling your task, because you can run it any time you need new performers.

Read more about <u>selecting</u> <u>performers</u> in our Knowledge Base.

atistics for 7 day bmitted tasks	Spent	Quality: control tasks	Quality: training tasks	Average submit time	Users	Banned users	
	0\$				0	0	
Active and close	ed Archived	Filters					Add trainir
		Filters Search	atic archiving applies to	pools with no activity	for 30 days).		Add trainir
Active and close Pools can be are Title \$		or automatically (autom		pools with no activity	-	ted \$	Add trainin
Pools can be arc	chived manually o	or automatically (autom	atic archiving applies to		Star		

2. Use the existing project instructions.

Instructions

✓ Use project instructions ⑦

[] • · ·

Website who sale goods for the people sometimes faced the problems with irrelevant products. In order to keep relevant information for customers your task is to classify product to the classes. So that we can update actual information on the website.

Your task is to classify how relevant is the product.

3.	Specify the training pool settings:	
<u>J</u> .	speeny the training poor settings.	

3.1. Click Create training.

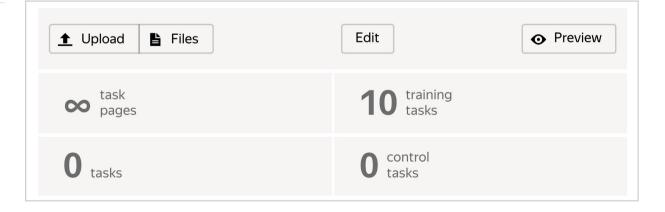
Read more about <u>training pools</u> in our Requester's Guide.

General settings	
Training title	Search relevance training $\qquad imes$
Price per task	🗹 free training
Adult content	No
Time per task suite	600 × seconds
Retry after	2 × days
 Assign in order of uplo Shuffle on page 	ading 💿
Settings for passir	ng training 💿
Full completion	Yes
Required to pass	Number of pages
Create training]

 Upload training tasks to the pool. You can download a sample file with tasks <u>here</u>: <u>Origin</u> License: PDDL

File upload settingsTasks per pageImage: Image: Image:

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



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 pool when there are no main tasks available anymore.

Search relevanc	e training — closed		Statistics	Edit	~
Download the sample file, add your task data The sample file uses TSV format, which is the Make sure you choose UTF-8 encoding when Template for general tasks.tsv Template for control tasks.tsv Template for training tasks.tsv	e same as CSV but using tab as the separator.				
<u>↑</u> Upload E Files	Edit O Preview	0 Users who completed training			
contask pages	10 training tasks				
O tasks	0 control tasks				

Create an exam pool

1. Click Add a pool.

Note: We recommend adding an exam pool along with the training because relevance evaluation projects are usually more complicated than most crowdsourcing projects, and it takes a certain effort to master all the guidelines. The more guidelines there are, the greater will be the need to check if the performers have really learned them.

itatistics for 7 days	s						
	= Spent	Quality: control tasks	Quality: training tasks	Average submit time	Users	Banned users	
(0\$			-	0	0	
ools Training	Statistics	Quality control					1
Active and closed	d Archived	Filters Search					Add a po
Pools can be arch	nived manually o	r automatically (autom	atic archiving applies to	pools with no activity	for 30 days).		
	Priority \$	Prog	ress	Status 🗘		Started \$	To be complete

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.

 POOL NAME (VISIBLE ONLY TO YOU)
 Classify search query relevance - exam

 Image: Classify search query relevance - exam
 Image: Classify search query relevance - exam

 Image: PUBLIC DESCRIPTION
 Image: Classify search query relevance - exam

 PUBLIC DESCRIPTION
 Image: Classify search query relevance - exam

 Image: PUBLIC DESCRIPTION
 Image: Classify search query

 Image: PUBLIC DESCRIPTION
 Image: Classify search query

 Image: PUBLIC DESCRIPTION
 Image: Public query

 Image: PUBLIC DESCRIPTION
 Image: Public query

 Image: Public query
 Image: Public quer

- 3. Specify pool parameters:
- 3.1. Select pool type. Choose Exam.

Note: The "Exam" type will allow you to launch these tasks with a price of zero. You can set a regular price, though. Setting a zero price might be reasonable if the exam is not very long but still complicated and you expect a low pass rate. That way you won't spend too much of your budget on the exam.

- 3.2. Attach the training pool you created earlier.

Note: The quality of the training can be low to just filter out potential deception, because we expect performers to make mistakes and learn from them (yet again, relevance is a complicated task type).

3.2			
	TRAINING	Search relevance training	~
	LEVEL REQUIRED, %	10 ×	

3.3. Set the price per task suite (for example, \$0.03).

Note: You can use a zero price as well. However, if the exam is time-consuming, a zero price might be unfair, as the performers will spend a lot of time completing it.

Read more about <u>pricing principles</u> in our Knowledge Base.

	Price per task suite		
	Each task suite can have one price for all tasks in the suite.		e page. Enter the total
PRICE IN US DOLLARS (?)	0.03 ×	FEE 📀	0.006
	+ Dynamic pricing	_	

3.4. <u>Filter</u> 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.

	Performers	Copy se	ettings from
	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		
ADULT CONTENT 2	No Add filter	~	Add skill

3.5. Choose Client = Toloka web version.

Then choose Languages = English. This way, performers who speak English will be invited to complete this task.

Client	→ = Toloka web version → 🗊 -
	AND
PERFORMER PROFILE	
Languages	✓ = English ③ ① -
	Performers who passed the language test ?
	language test

3.6. Set up exam quality calculation.

Click Add skill.

	Performers	Copy s	ettings from
ADULT CONTENT 📀	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	3.6	
	Add filter	~	Add skill

3.7. Fill in the skill title. Add a description if necessary. Click **Add**.

TITLE	
search_relevance_exam	×
DESCRIPTION	
Public? No 2	
	3.7

3.8. Scroll down to **Quality Control**. Add a **Control tasks** rule.

> We will have 10 tasks in the exam pool, so the quality will be calculated after the whole exam has been passed.

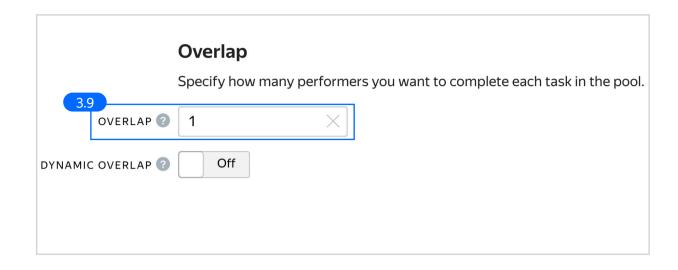
Note: We will then use this parameter as an entry filter for the main pool.

CONTROL TASKS 😨	
Recent control task responses to use items	
3.8 If number of responses \sim \geq 10 \times +	
then assign skill from the field v search_relevanc ×	
correct responses (%) V	+

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

Note: If you are using the **"Exam"** task type, the overlap will be infinite by default, so you don't need to change that field.

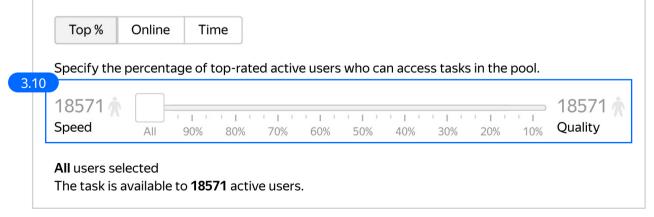
To understand <u>how this rule works</u>, go to the Requester's Guide.



3.10. As we are assessing people in this pool the **Speed/quality** balance can be left as default.

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...



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

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

1	Parameters				
	TIME PER TASK SUITE IN 3 600	\times	POOL CLOSING DATE 🗿	2022-09-16	
	KEEP TASK ORDER ? No		WAITING TIME FOR THE POOL TO CLOSE IN (?) SECONDS	0	
		PO	OL PRIORITY WITHIN THE PROJECT	0	

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

	Parameters		
TIME PER TASK SUITE IN SECONDS	600	POOL CLOSING DATE ?	2022-09-23
KEEP TASK ORDER 🕢	No	WAITING TIME FOR THE POOL TO CLOSE IN ? SECONDS	0
		POOL PRIORITY WITHIN THE PROJECT	0
4- Cancel	Save		

5. Upload exam tasks to the pool. You File upload settings 💿 can download a sample file with exam tasks here: Origin Tasks per page License: PDDL Note. You can create the exam tasks locally and upload them in a file like this one, or upload regular By empty row Set manually Smart mixing tasks and create controls via the UL Main tasks 0 Read more about this in the Training tasks 0 Requester's Guide. Control tasks 10 \times Set 10 tasks per page. Show advanced settings Upload Sample file for uploading tasks Close

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

▲ Upload Files Delete	Edit	• Preview
task pages	0 training tasks	
O tasks	10 control tasks	

7. Run the exam pool.

Note: We recommend opening the exam pool along with the main pool or just before. Otherwise Tolokers will spend their time on the exam pool but get no access to real tasks and might not return to the project later.

Classify search o	query relevance - exam	— closed	Statistics 🖢 Download results 🗸	Edit 🗸
Download the sample file, add your task data, The sample file uses TSV format, which is the Make sure you choose UTF-8 encoding when	same as CSV but using tab as the separator.			
Template for general tasks.tsv				
Template for control tasks.tsv				
Template for training tasks.tsv				
Lupload B Files B Delete	Edit • Preview		0 %	
coo task pages	0 training tasks		Completed 0	
O tasks	10 control tasks	0		00

Create the main pool

1. Click Add a pool.

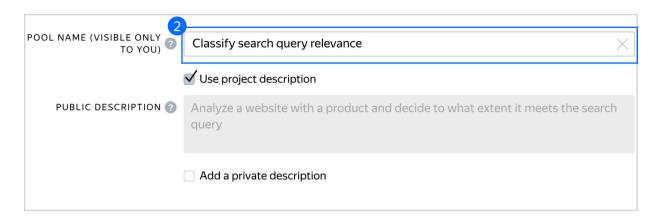
A pool is a set of paid tasks grouped into task pages. These tasks are sent out for completion at the same time.

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

ubmitted tasks Sr	pent Ouality:	control tasks Quality: training	tasks Average submit time	Users B	anned users	
	\$		-	00		
Pools can be archi	ived manually or automat	tically (automatic archiving ap	plies to pools with no activity	for 30 days).		
Title \$	Priority \$	Progress	Status 🔶		arted ≑	To be complete

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.



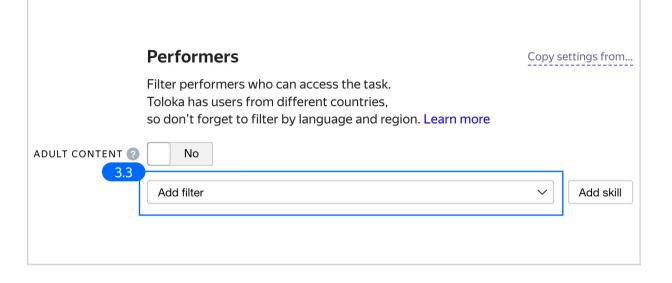
- 3. Specify pool parameters:
- 3.1. Select pool type. Choose Other.

Read more about <u>pool types</u> in the Requester's Guide.

POOL TYPE ?	Other	^	
	_		
TRAINING	Exam		\checkmark
	Training		
3.1	Retry		
	✓ Other		
	Lacii lask suile caii ia	IVE ONE	or multiple tasks on the same page. Enter the total

3.2. Set the price per task suite (for example, \$0.03).		Price per task sui	ite	
Read more about <u>pricing principles</u> in our Knowledge Base.		Each task suite can have one or multiple tasks on the same page. Enter the top price for all tasks in the suite.		
		+ Dynamic pricing		

3.3. <u>Filter</u> performers who can access the task. Choose "**No**" in the **Adult content** block.



3.4. Add a language filter and limit the access to PC users only. The task involves a lot of actions (examining the image, analyzing search results for the query) which are not very mobile-friendly.

CALCULATED DATA					
Client	~	= Toloka web versio	n 🗸	Û	+
	A	ND			
PERFORMER PROFILE		~			
Languages	\sim	= English	8	Û	+
		Performers who passe	ed the		

3.5. Add filtering by exam quality. Select **My skills** and find a skill where you recorded exam quality. Use it as a filter. The higher the access threshold is, the better the quality will be.

						Skills
CALCULATED DATA	~	= Toloka web version	~	Ŵ	+	My skills Choose global skill
PERFORMER PROFILE		ND = English	۲	Û	+	Skill from " Ya.Hamilto Skill from " Agros"
	A	Performers who passer language test ?	d the			Skill from " Alverca" Skill from " Amaz"
skills search_relevance_exa	m 🙁	≥ 90		ŵ	+	

3.6. Set up <u>Quality control</u>.

Read more about <u>quality control</u> <u>principles</u> in our Knowledge Base or check out <u>control tasks settings</u> in the Requester's Guide.

Add rules to get more accurate responses. All rules work independently.					
CAPTCHA FREQUENCY ② None ~		_			
3.6	ON-AUTOMATIC ACCEPTANCE	No	REVIEW PERIOD IN DAYS	14	
			CAPTCHA FREQUENCY	None	~
	3.	6	+ Add Quality Control F	Rule	

Quality control

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

Note: The skill will start calculating after we get at least 3 responses, and will use a history of the 10 last responses. Based on this history, we will record a quality parameter in the same skill we used as a quality filter. If you want to record exam results and current quality in different skills, create a new one, use it for this rule, and also use it as a filter.

3.8. Set up the up the <u>Fast responses rule</u>.

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

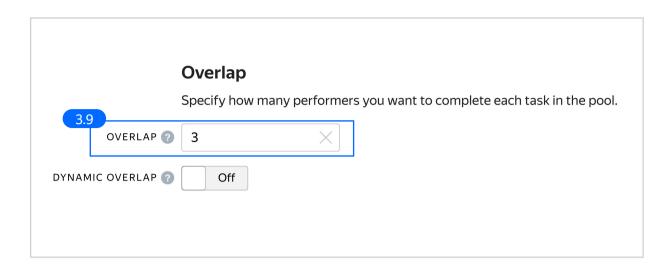
Note: In this configuration, a performer will be banned for two days if they give an answer within less than 10 seconds.

3.7 Recent control task responses to use 20 ×	
If number of responses \checkmark \geq 1 \times +	
then assign skill from the field \checkmark search_relevanc \times	
correct responses (%)	÷

ST RESPC	DNSES 3		
Recent	task suites to use items		
Minim	um time per task suite 10 \times		
8 If	number of responses V ≥ 1 × +	_	
and	number of fast responses \checkmark = 1 \times = +		
then	ban v on project v		
	2 × days v		
	Fast resonse	×	

3.9. Set an overlap. This is the number of users who will complete the same task. For example, 3. We will aggregate the results after the pool is completed.

To understand <u>how this rule works</u>, go to the Requester's Guide.



3.10. Optionally, specify the percentage of top-rated performers in the <u>Speed / quality ratio</u>. This is not really necessary though, since we only grant access to performers who already proved their competence in the exam.

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...



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



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

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

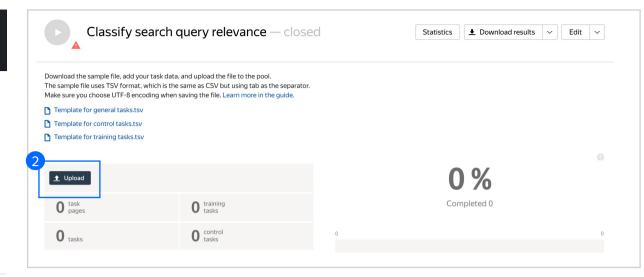
1	Parameters				
	TIME PER TASK SUITE IN 3 600	\times	POOL CLOSING DATE 🗿	2022-09-16	
	KEEP TASK ORDER ? No		WAITING TIME FOR THE POOL TO CLOSE IN 7 SECONDS	0	
		PO	OL PRIORITY WITHIN THE ? PROJECT	0	

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

Parameters		
TIME PER TASK SUITE IN ③ 600	POOL CLOSING DATE (?)	2022-09-23
KEEP TASK ORDER ? No	WAITING TIME FOR THE POOL TO CLOSE IN ? SECONDS	0
	POOL PRIORITY WITHIN THE PROJECT	0
Cancel Save		

Prepare and upload a file with tasks

- Prepare a TSV file with tasks as shown in our <u>example</u>. <u>Origin</u> License: PDDL
- 2. <u>Upload pool tasks</u> from this file.



2.1. Select <u>Smart mixing in File upload</u> settings and specify the number of tasks of each type per page.

> Note: 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 <u>grouping tasks</u> into suites, read the Requester's Guide.

Click Upload again.

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

File upload settings			
Tasks per page			
By empty row	Set manually		Smart mixing
Main tasks		4	×
Training tasks		0	
Control tasks		1	×
Show advanced settings			
Sample file for uploading	tasks		2.1 Close Upload

3. Create control tasks.

Click Edit \rightarrow Create control tasks.

Classify search o	query relevance — close	d	Statistics	Download results	~	Edit	~
Download the sample file, add your task data The sample file uses TSV format, which is the Make sure you choose UTF-8 encoding when Template for general tasks.tsv Template for control tasks.tsv Template for training tasks.tsv	same as CSV but using tab as the separator.						
1 Upload E Files 2 Delete	Edit		0	%			
0 task pages	0 training tasks		Com	pleted 0			
123 tasks	0 control tasks	0					0

 $\label{eq:Projects} Projects \rightarrow Classify \, \text{search query relevance} \rightarrow Classify \, \text{search query relevance} \rightarrow Uploaded \, \text{tasks}$

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

Aain 113 Control t	asks 10 Training	tasks 0	
Create control tasks	Create training tas	ks	Download
ID \$	Overlap 🗍 🛛	Responses from performers	Last updated 👙
07b7a53c	3	0	09/22/2021 12:54:06 PM
07b7a541	3	0	09/22/2021 12:54:06 PM
07b7a537	3	0	09/22/2021 12:54:06 PM

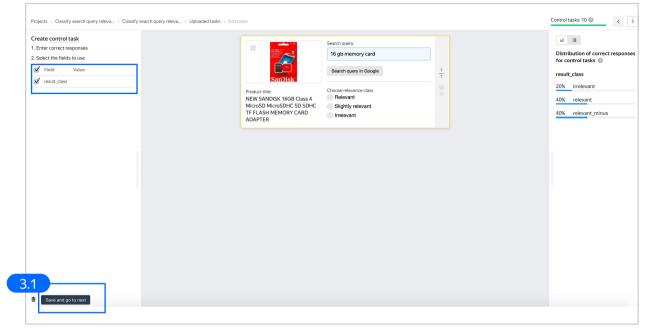
Note: Control tasks are tasks that already contain the correct response. They are used for checking the quality of responses from performers. The performer's response is compared to the response you provided. If they match, it means the performer answered correctly. 3.1. Enter correct responses to your control tasks. Check the **result_class** 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 ~10 control tasks.

Note: In small pools, control tasks should account for 10-20% of all tasks.

To learn more about <u>creating control</u> <u>tasks</u>, go to the Requester's Guide.

Note: You can check the distribution of labels for control tasks in the right panel. We recommend making it even to ensure a fair quality calculation.

Note: Control tasks should be as reliable as possible. It may be a good idea to evaluate control tasks within your team to ensure the guidelines are really clear about these cases, and that all details about the query have been considered.



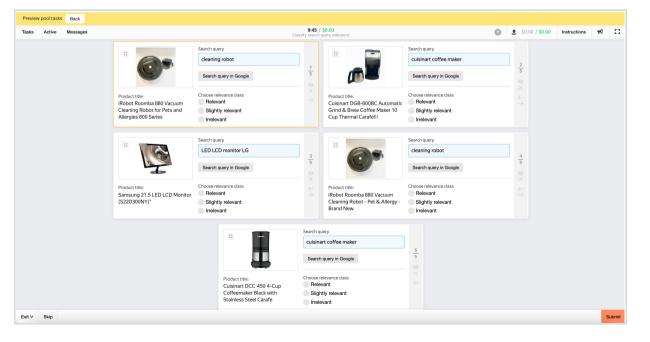
3.2. To check the number of control tasks, go to the **Pool page.**

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 Origin License: PDDL

▲ Upload Files Delete	Edit	• Preview
~87 task pages	0 training tasks	
113 tasks	10 control tasks	

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.



5. Start the pool.

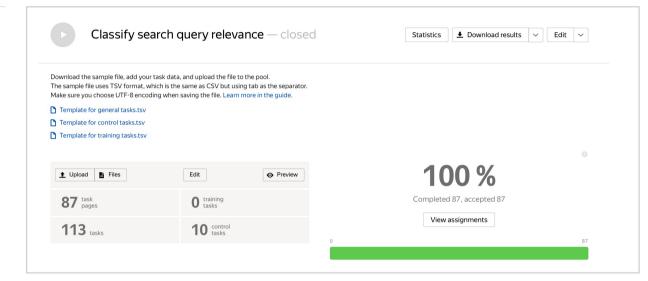
Make sure you start the training pool and the exam pool before you start the main pool.

	query relevance — clos	sed	Statistics J Download rest	ilts 🗸 Ed
Download the sample file, add your task dat The sample file uses TSV format, which is th Make sure you choose UTF-8 encoding whe	e same as CSV but using tab as the separate	or.		
Template for general tasks.tsv				
Template for control tasks.tsv				
Template for training tasks.tsv				
1 Upload B Files Delete	Edit • Preview		0 %	
			• / •	
			Completed 0	
~87 task pages	0 training tasks			
~87 task pages	0 tasks	0		

Receiving responses

- 1. Wait until the pool is completed. Refresh the pool page to check progress.
- 2. Click Download results.

Note: Do not forget to close the training pool and the exam pool when there are no main tasks available anymore.



3. Make sure to uncheck **Separate** assignments with empty row and Download the results.

Download results			
Status	Active	Submitted	Accepted
	Rejected	Skipped	Expired
Columns	VRL	🗹 task ID	🗹 assignment ID
	Task suite ID	🗹 Performer ID	✓ status
	🗹 start time	submit time	accept time
	reject time	Skip time	expire time
	price		
Download data for the period			
Separate as	signments with empty row		
Exclude ass	ignments by banned users		3
			Close Download results

Appendix

Interface code Step 4.1. { "view": { "type": "view.list", "items": [{ {
 "type": "layout.sidebar",
 "minWidth": 400,
 "content": {
 "type": "view.list",
 "size": "m",
 "direction": "vertical",
 "items": [
 {
 {
 } "type": "view.image", "maxWidth": 350, "url": { "type": "data.input", "path": "imagepath" {
 "type": "view.markdown",
 "label": "Product title:",
 "content": {
 "type": "data.input",
 "path": "title" }, "controls": { "type": "view.list", "direction": "vertical", "items": [{ { {
 "type": "view.alert",
 "theme": "info",
 "label": "Search query",
 "content": {
 "type": "view.text",
 "content": {
 "type": "data.input",
 "path": "query"
 }
} "type": "view.action-button", "label": "Search query in Google", "action": { "type": "action.open-link", "payload": { "type": "data.input", "path": "search_url" "type": "view.divider" "type": "field.radio-group", "label": "Choose relevance class", "options": ["label": "Relevant", "value": "relevant" "label": "Slightly relevant", "value": "relevant_minus" { "label": "Irrelevant", "value": "irrelevant"], "data": { "type": "data.output", "path": "result_class" "validation": {
 "type": "condition.required" "plugins": [{
 "type": "plugin.toloka",
 "layout": {
 "kind": "scroll", "taskWidth": 600