Vandex

Audio transcription project

Alexey Drutsa, Dmitry Ustalov, Valentina Fedorova, Olga Megorskaya, Daria Baidakova

Project setup

1. Click the button **+ Create project.**

2. Choose a **Blank** template. Blank Blank

r.

Projects Active V Only mine V	Find project + Create proj
类 Toloka Projects Users Skills Profile Messages	
For a quick way to get tasks set up and running, use templates. You can choose one of the standard templates, or adapt a template to your gover adat and response format. See the	
Toloka use cases in Requester's Guide for step-by-step instructions for launching projects.	
Toloka use cases in Requester's Guide for step-by-step instructions for launching projects.	

3. Fill in a title and a description of your project, then press **Save**.

General information		
Name to show performers *	***	**
Transcribe the recording	Trans	cribe the recording
Description for performers	Туре и	hat you hear in the recording.
Type what you hear in the recording.	0\$ per tas	∼0\$ k per hour
Private comment 🔕		

4. Make sure "Template builder" is chosen. Delete the existing config and paste <u>the new</u> <u>code from here</u>.



5. Press on **Input Data example** in the top right corner.

6. Paste <u>the following code</u> for Input Data Example so it looks like on the picture.

Θ 2 Task interface Editor HTML / JS / CSS ② O Template builder ③ C 0 Input data example Config Input data example "vars": { "author": "assessorlead", "audio": "https://yastat.net/s3/tb/static/file "task": "new", } "validString": "type": "condition.all", "type": "condition.sc 9 10 "type": "string", "pattern": "^[a-zA-12 13 14 15 16 17 18 19 20 "hint": "Use only lowerca Transcription text "view": { "type": "view.list", 21 "items": [23 "type": "view.audio", 24 25 26 "type": "condition.pl "hint": "You didn't l

7. Make sure the specifications look like this and save the template:

Output data	<>
result_unformat (string)	
result (json)	
Add field	
	result_unformat (string) result (json)

8. Write an instruction or just paste <u>our</u> <u>version</u>. Then **Save** this block.

9. Leave the "Translations" section as it is

When a performer selects a task, they first se	ee the instructions that you wrote. Describe what needs to be done and give examples.	
You can prepare your instructions in HTML for	ormat, then copy and paste them into the editor. Press < > to switch to HTML mode. To le	earn more, see
the documentation.		
-T (₽) B I U 🖪 ⇔ 🗄 🗄 🗄 🗄		[] • •
The training has 7 tasks. Upon successful co	mpletion the main pool will open.	
Each page includes several tasks with short	audio recordings.	
Listen to them and type the text you hear.		
Use headphones to hear the speech more cl	early.	
Please, listen carefully and try to replay and	re-listen the recording in problem cases. When you finish transcribing, carefully reread t	the text.
Use a dictionary if you are not sure of word s	spelling correctness.	
Basic rules:		
 use only lowercase letters (for example 	e. "mv cat is awesome" instead of "Mv cat is awesome")	
Save		
Save Translations		
Save Translations		
Save Translations Performers from different countries	s will understand the purpose of the task better if the instructions and all descriptions a	ire in their
Save Translations Performers from different countries native language. Fill in "Name and c	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If yo	ire in their u leave these
Save Translations Performers from different countries native language. Fill in "Name and c empty, the language is inactive.	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If yo	ire in their u leave these
Save Translations Performers from different countries native language. Fill in "Name and c empty, the language is inactive. Source language	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If yo	are in their u leave these
Save Translations Performers from different countries native language. Fill in "Name and c empty, the language is inactive. Source language	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If yo	ire in their u leave these
Save Translations Performers from different countries native language. Fill in "Name and c empty, the language is inactive. Source language _	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If yo	are in their u leave these
Save Translations Performers from different countries native language. Fill in "Name and c empty, the language is inactive. Source language _	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If yo	are in their u leave these
Save Translations Performers from different countries native language. Fill in "Name and c empty, the language is inactive. Source language _ Translations	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If yo	ire in their u leave these
Save Translations Performers from different countries native language. Fill in "Name and c empty, the language is inactive. Source language - Translations Language Language	s will understand the purpose of the task better if the instructions and all descriptions a lescription" and "Instructions" for each language that you want performers to see. If you	ire in their u leave these

Add translation

Save 🗸

	10	. Save	the	proj	ect
--	----	--------	-----	------	-----

and **Save** this block.

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

Training pool creation

1. Go to the "**Training**" tab and click the "**Add training**" button.



 Keep the project instructions and fill in pool settings. Then click the "Save training" button.

Read more about training pools in the Toloka Requester's Guide.

General settings	
Training title	train_naaci
Price per task	Stree training
Adult content	No
Time on task	1200 × seconds
Retry after	5 × days
Assign in order of up Shuffle on page	oloading 💿
Settings for passi	ing training 💿
Full completion	Yes
Required to pass	Number of pages
Save training	

3. Upload training tasks to the pool. You can download the file with the tasks here: https://tlk.s3.yandex.net/naacl2021/data_train.tsy



4. If the upload went fine, you will see the task counter updated. PRESS PLAY



5. Run training pool by pressing large blue button.



Main pool creation

1. Go to the "**Pools**" tab and click the "**Add a pool**" button.

Toloka Projects Users Skills Profile Messages	
Transcribe the recording — active ★ ★ ★ ★ ★ 4.23. 12 votes ◎	Project actions \vee
Statistics for 7 days Quality: control tasks Quality: training tasks	Average submit time Users Banned users
0 0 \$	0 0
Active and closed Archived Filters Search Pools can be archived manually or automatically (automatic archiving applies to	Add a pool

2. Fill in the pool settings (see more on the next page).

OL NAME (VISIBLE ONLY TO YOU)	Transcribe the recording	×
	Use project description	add a title
PUBLIC DESCRIPTION 🕖	Type what you hear in the recording.	
	Add a private description	
POOL TYPE 🚷	- ~	attach your training pool
TRAINING	train_naacl	~
LEVEL REQUIRED	85 ×	select your training quality threshold



4. Add basic performer filters.



F	Performers			Copy se	ttings from	
F T s	ilter performers who ca oloka has users from di o don't forget to filter b	an access the ta ifferent countrie by language and	sk. es, Fregion. Learn mor	e		
ADULT CONTENT 📀	Yes		 the task will on agreed to work 	ly be visible f with adult co	or those who ntent	
English is required	Add filter		~ C	reate a skill		
-	PERFORMER PROFILE					
~	Languages	~ =	English	×	+	
		AND				
	CALCULATED DATA		1			
	Client	~ =	Toloka web version		+	
-		OR =	Toloka for mobile	*		

5. Set up **Control tasks** quality control rule.

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 5 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 <u>quality control rules</u>.

6. Set up Fast responses quality control rule.

NTROL TA	control task responses to use items
lf and	number of responses 2 3 × + correct responses (%) < 60 × + < 60 × = + < 60 × = + × < 60 × = + + + + + × × < 60 × = + + +
then	ban Image: mail on project 5 X days Image: mail of the project
	control tasks ×

FAST RESPC	NSES 2		
Recent	task suites to use 10 \times um time per task suite 30 \times		
lf then	number of fast responses > ban > on project > 5 X days > fast responses		

7. Select overlap **3** for each task.

We will aggregate the results after we finish our pools.

8. Set up a rating threshold — the percentage of top-rated users who will see your task.

9. Fill in time allowed per task page completion.

	Overlap			
	Specify how many performers you want to complete each task in the pool.			
OVERLAP 🕐	3	\times		
DYNAMIC OVERLAP 🥐	Off			
Speed/quality ratio	erformer access based on th	eir rating in Toloka.		
This boosts quality but may slow fewer performers available to co	down project completion be implete tasks. Learn more	ecause there will be		
Top % Online Time				
Specify the percentage of top-rated	ictive users who can access tasks	in the pool.		
7922 A	1 60% 50% 40% 30%	4753 m 20% 10% Quality		
60% top-rated performers were select The task is available to 4753 active us	ted. Jers.			
Pa	arameters			
TIME PER TASK SUITE IN @	00 ×	POOL CLOSING DATE 😨	2022-05-23	
TIME PER TASK SUITE IN @ SECONDS @ SKEEP TASK ORDER @	No	POOL CLOSING DATE ② WAITING TIME FOR THE POOL TO CLOSE IN ③ SECONDS	2022-05-23 0	111

10. Click on Upload button and choose **Smart mixing** option for your task upload.



Tasks per page		
By empty row Set manu	S Smar	t mixing
Main tasks	4	×
Training tasks	0	
Control tasks	1	\times
Show advanced settings		
Sample file for uploading tasks	Close	Upload

11. Leave the settings in place and upload the dataset. You can find it here: https://tlk.s3.yandex.net/naacl2021/data main.tsv

	Preview pool tasks Back		
	Tasks Active Messages	14:44 / \$0.03 Transcribe the recording S \$0.00 / \$0.00 Instructions	τ0 II
12. Preview the pool and make sure the tasks submit fine.	Q ► 0.00/0.03		
	Transcription text	$\xrightarrow{\times}$	
	Q > 0.00/0.06	→ 4) : $\frac{2}{5}$	
	Transcription text		
	Q > 0.00/0.04	▲ ● : 3 5 ⊙	
	Transcription text		
	Q > 0.00/0.04	•) : ⁴ / ₅	
	Exit Y Skip		Submit

13. Run the main and the training pool.



14. Once the labeling is completed, you can download the results by clicking on the **Download Results** button.

Make sure to untick the box "Separate assignments with empty row".



15. Aggregate the results by using Text Aggregation. You can <u>find the code in Toloka</u> <u>Crowd-Kit</u>. Please, make sure that you download the results and name the file <u>assignments.tsv</u>

Useful resources

Tutorial slides

https://research.yandex.com/tutorials/crowd/naacl-2021

Training Pool Dataset

https://tlk.s3.yandex.net/naacl2021/data_train.tsv

Main Pool Dataset

https://tlk.s3.yandex.net/naacl2021/data_main.tsv

Text aggregation using Crowd-Kit

https://tlk.s3.yandex.net/naacl2021/HRRASA.ipynb

Toloka

https://toloka.ai/