Toloka

Toloka

ICWE 2022 hands-on tutorial

Web Engineering with Human-in-the-Loop

Dmitry Ustalov, Nikita Pavlichenko, Boris Tseytlin,
Daria Baidakova and Alexey Drutsa

Part IV
Practice Session

Daria Baidakova,
Director of Educational Programs

Tutorial Schedule

Part I Intro — 45 min
The Role of HITL in building Search Engines

Part II — 45 min
Ranking and Quality Metrics

Coffee Break — 30 min

Part III — 45 min
Human-in-the-Loop Essentials

Part IV — 45 min
Practice: Websites relevance

Lunch Break — 90 min

Part V — 45 min
Results aggregation and implementation into ML pipeline

Part VI — 90 min
Metric Computation

Coffee Break — 30 min

Part VII — 90 min
Results discussion and conclusion

Practice session: scope

Imagine that you develop a machine learning pipeline to help improve search quality

- You have a dataset of queries and search results (websites) for each query
- You need to get a ranked list of search results for each query
- These collected data will further be used to train a search algorithm

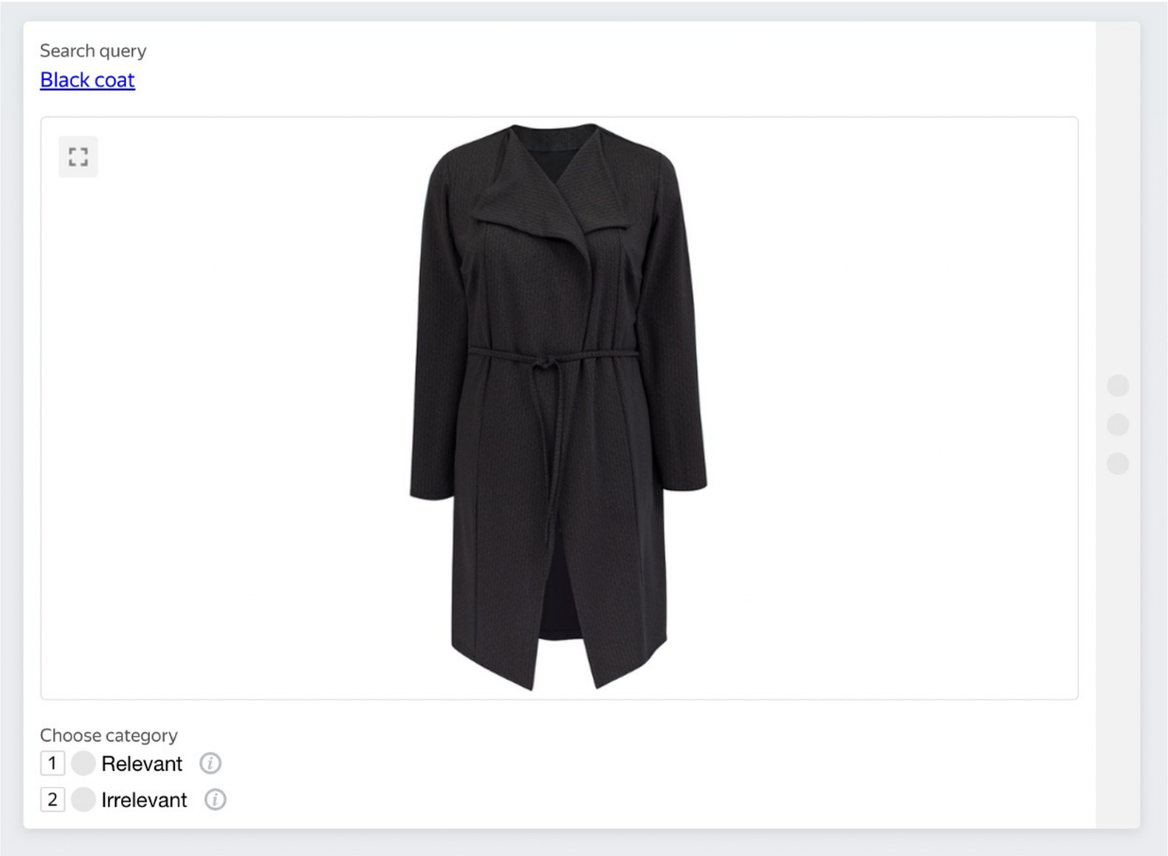
Practice session: scope

Imagine that you develop a machine learning pipeline to help improve search quality

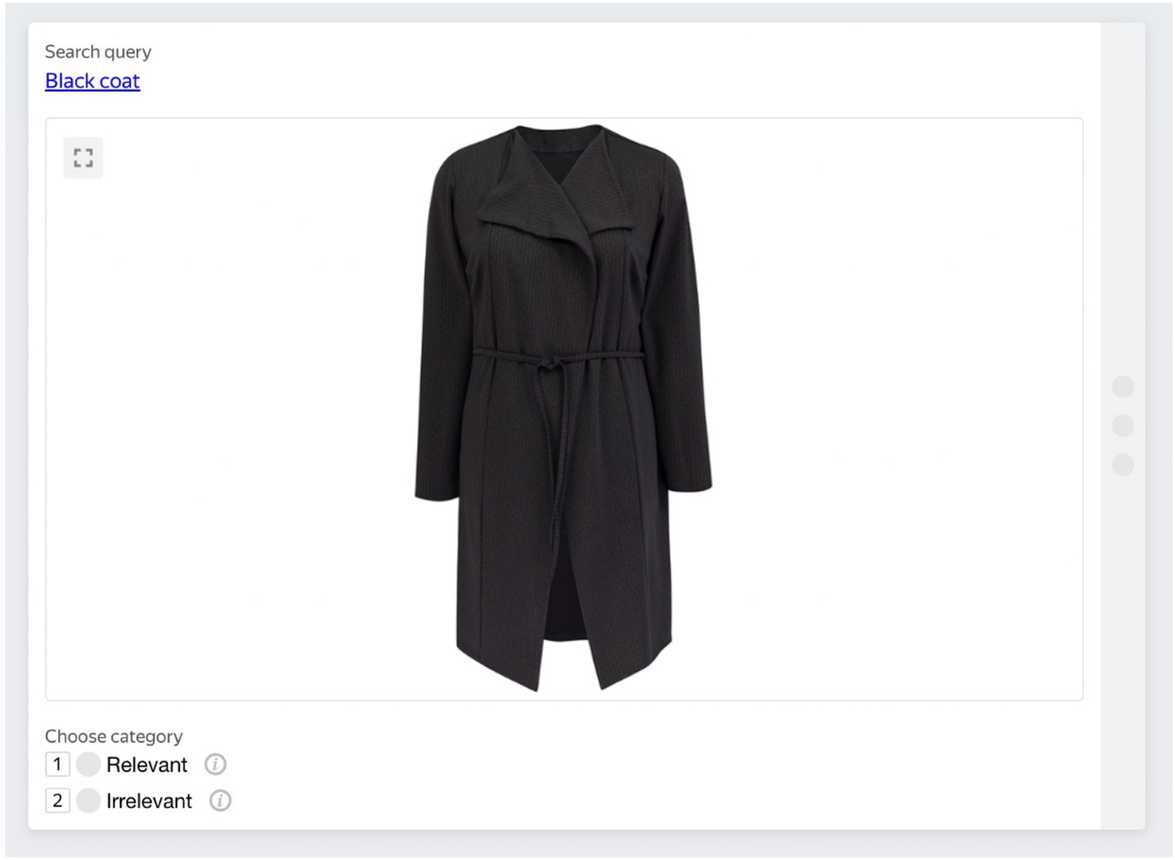
- You have a dataset of queries and search results (websites) for each query
- You need to get a ranked list of search results for each query
- These collected data will further be used to train a search algorithm

It is your goal during the practice session of our tutorial

Project: Is the webpage relevant to the query?



Project: Is the webpage relevant to the query?



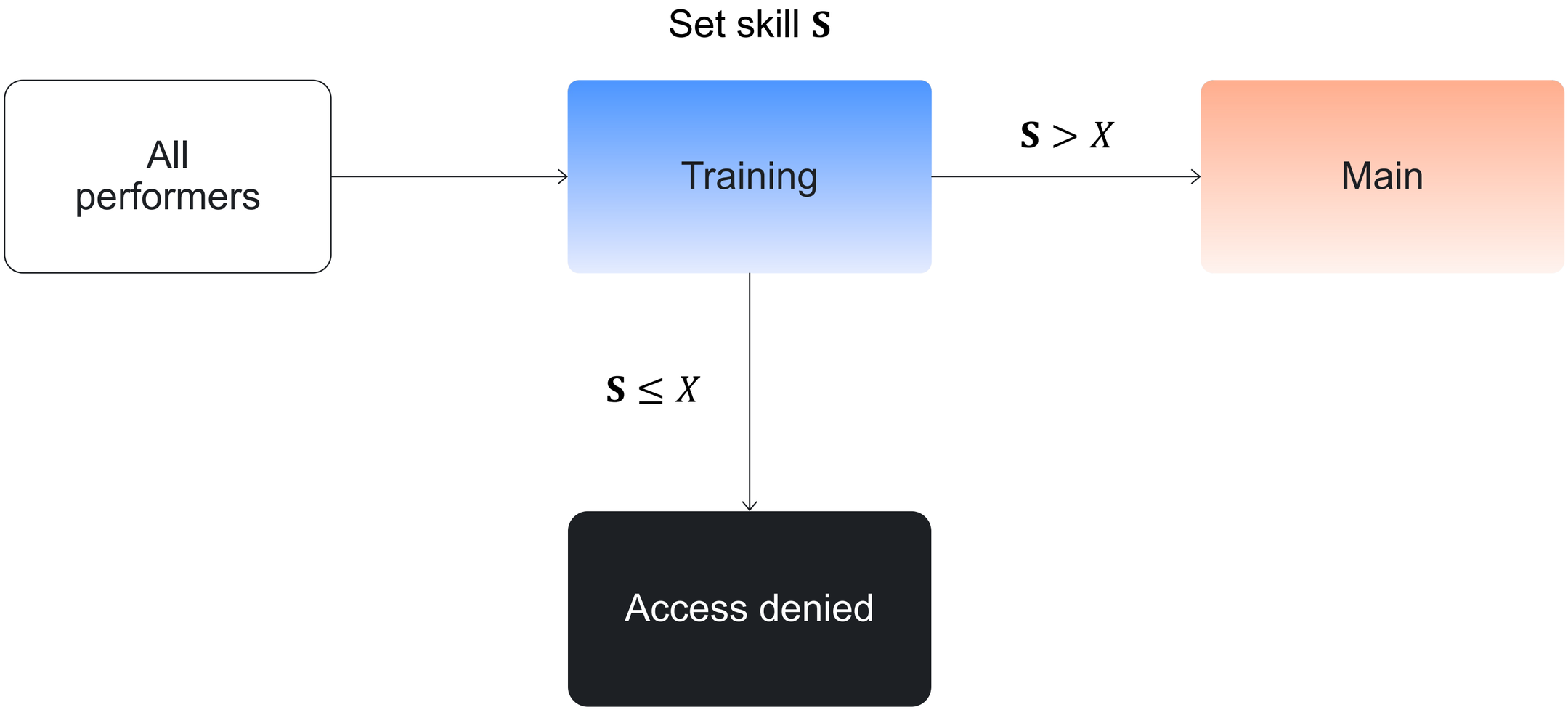
Task

Is the webpage relevant to the query?

Settings

- Type: Classification
- Training: yes
- Quality control: golden set
- Overlap: 3
- Pay: $0.05 per 1 page with 10 tasks
- Time: 600 sec per 1 page

Performer training



We are starting the practice session: 45 min

We share you links to:

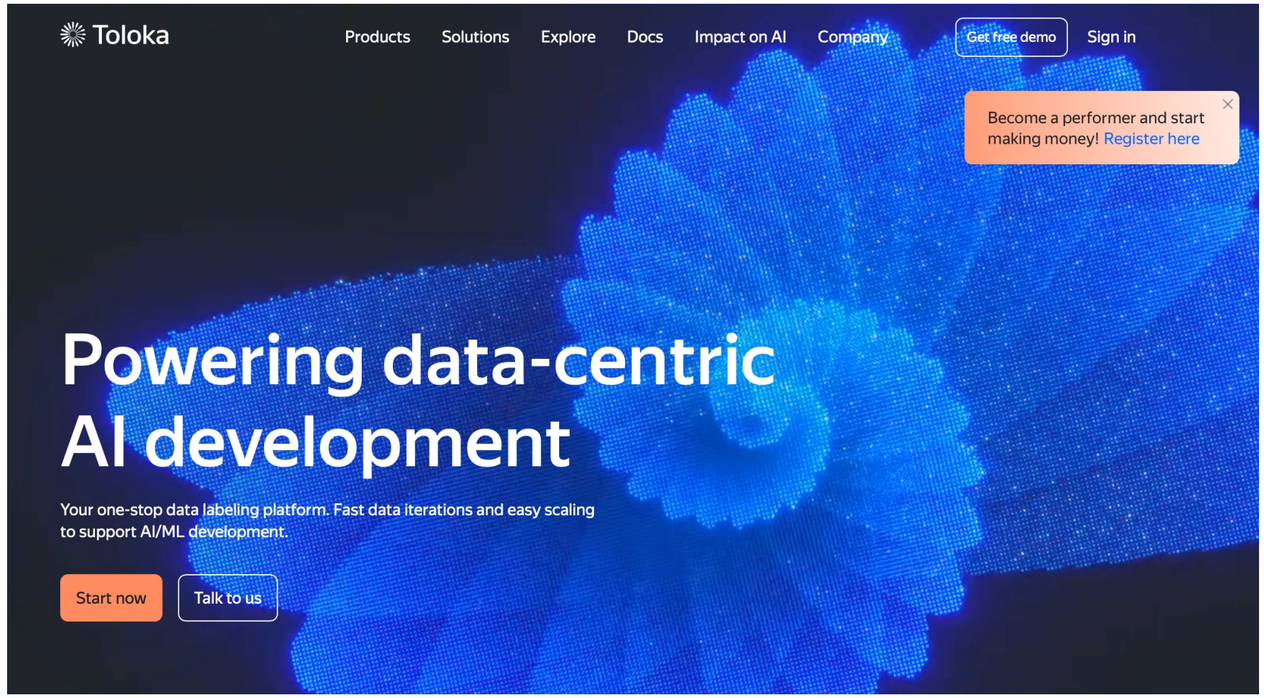
- A step-by-step guide to configure and run data labeling
- A dataset with queries and webpages
- Promo code to use Toloka as a requester

Step-by-step instruction:

https://toloka.ai/files/icwe_2022/guide.pdf

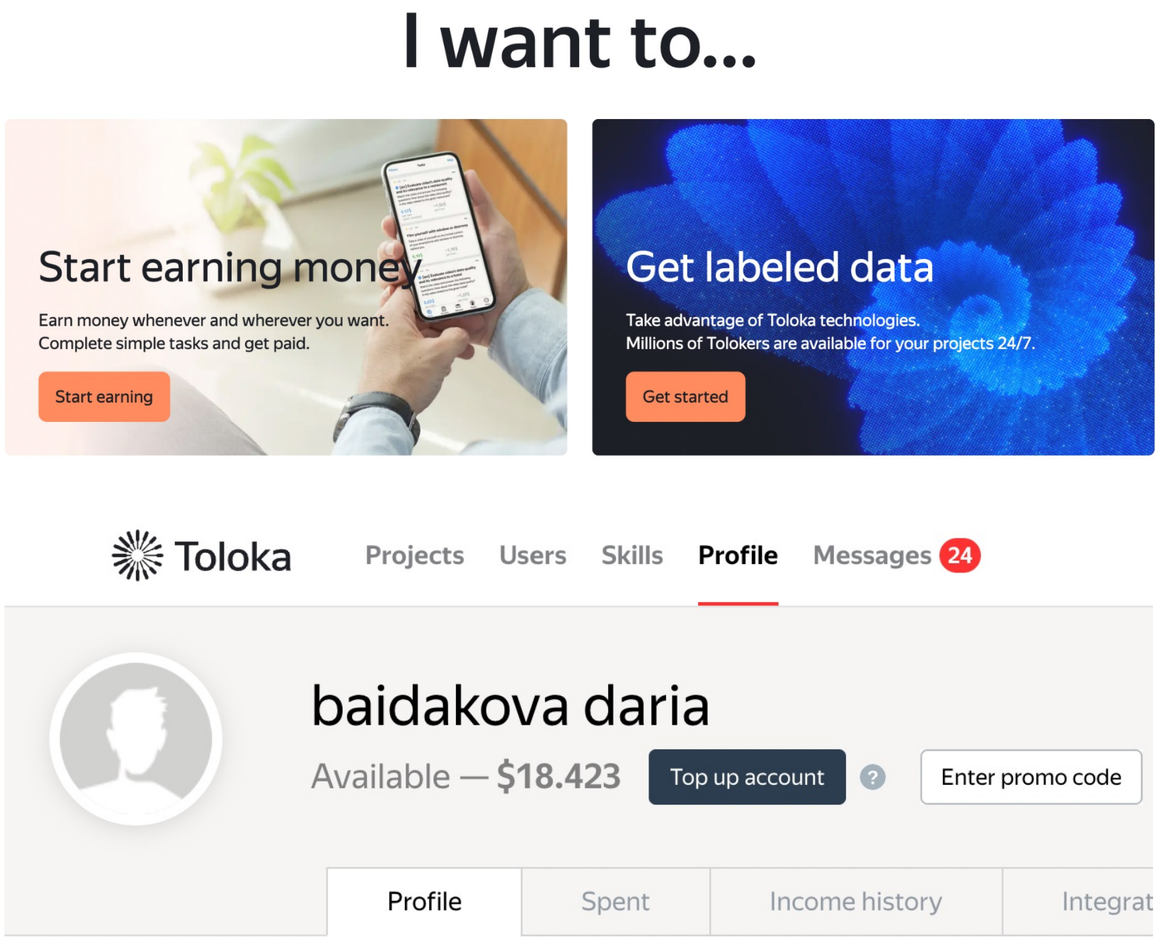
Sign up on Toloka as a requester

1. Go to toloka.ai
2. Click on “Start now”



Sign up on Toloka as a requester

3. Click on “Get labeled data”
4. Use your email and password
5. Go to Profile and use promocode



PROMO CODE: TOLOKA_ICWE2022

Toloka demo
5 min

Join our Slack: icwe_tutorial channel

Daria Baidakova

Director of Educational Programs

dbaidakova@toloka.ai

https://toloka.ai/events/icwe-2022/

Useful links

1. Config https://toloka.ai/files/icwe_2022/tb_config.txt
2. Input data https://toloka.ai/files/icwe_2022/tb_input_data.txt
3. Instructions https://toloka.ai/files/icwe_2022/instructions.txt
4. Train dataset https://toloka.ai/files/icwe_2022/train_dataset.tsv
5. Main dataset https://toloka.ai/files/icwe_2022/main_dataset.tsv