



Toloka

Improving Web Ranking with Human-in-the-Loop: Methodology, Scalability, Evaluation

Alexey Drutsa, Dmitry Ustalov, Nikita Popov,
Daria Baidakova

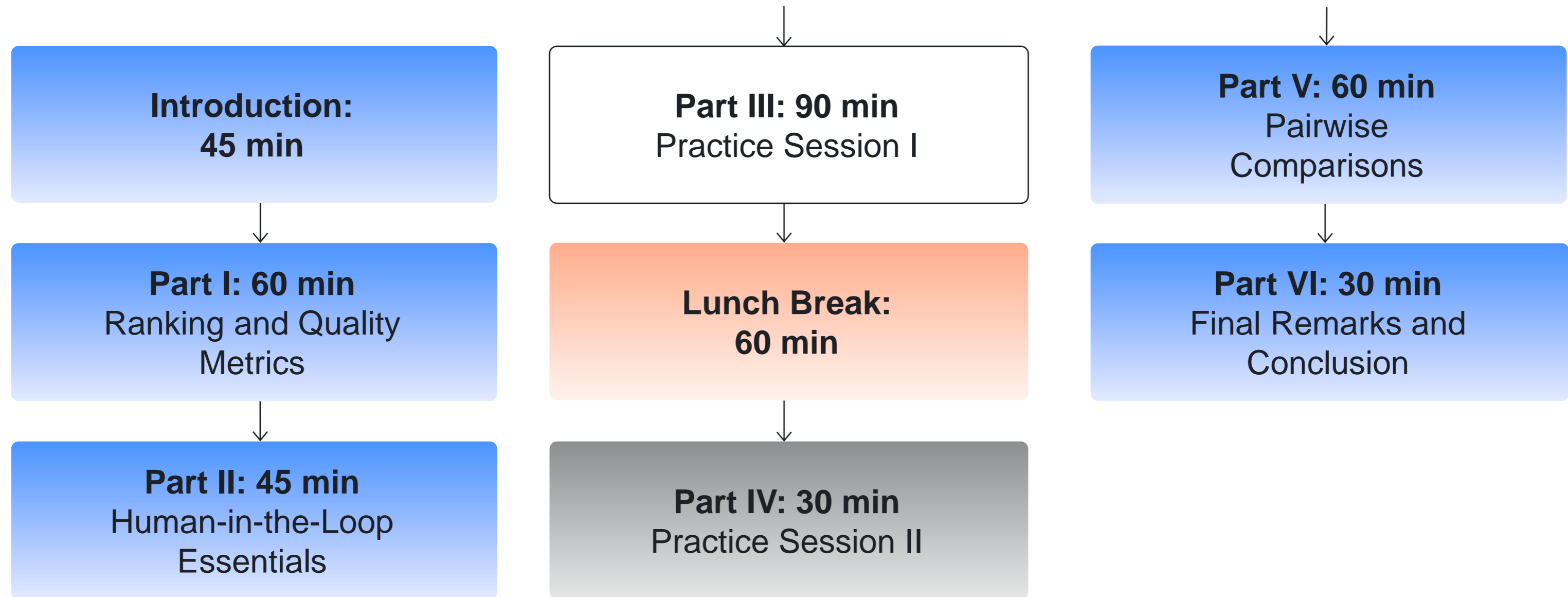


Part III

Setting up and running label collection projects

Daria Baidakova,
Sr. Manager, Customer Success & Education

Tutorial schedule



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 (URLs) 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

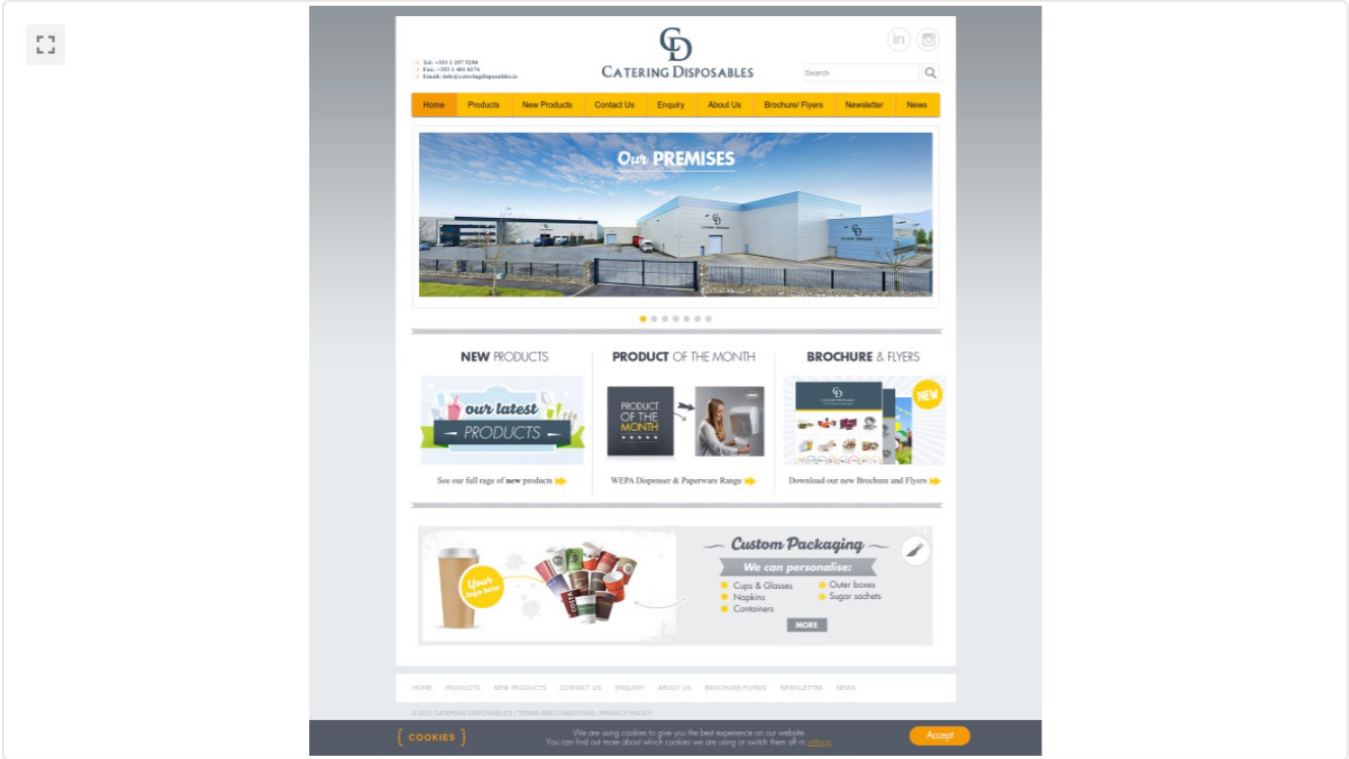


Project: Which of the two sites more accurately matches the search query?

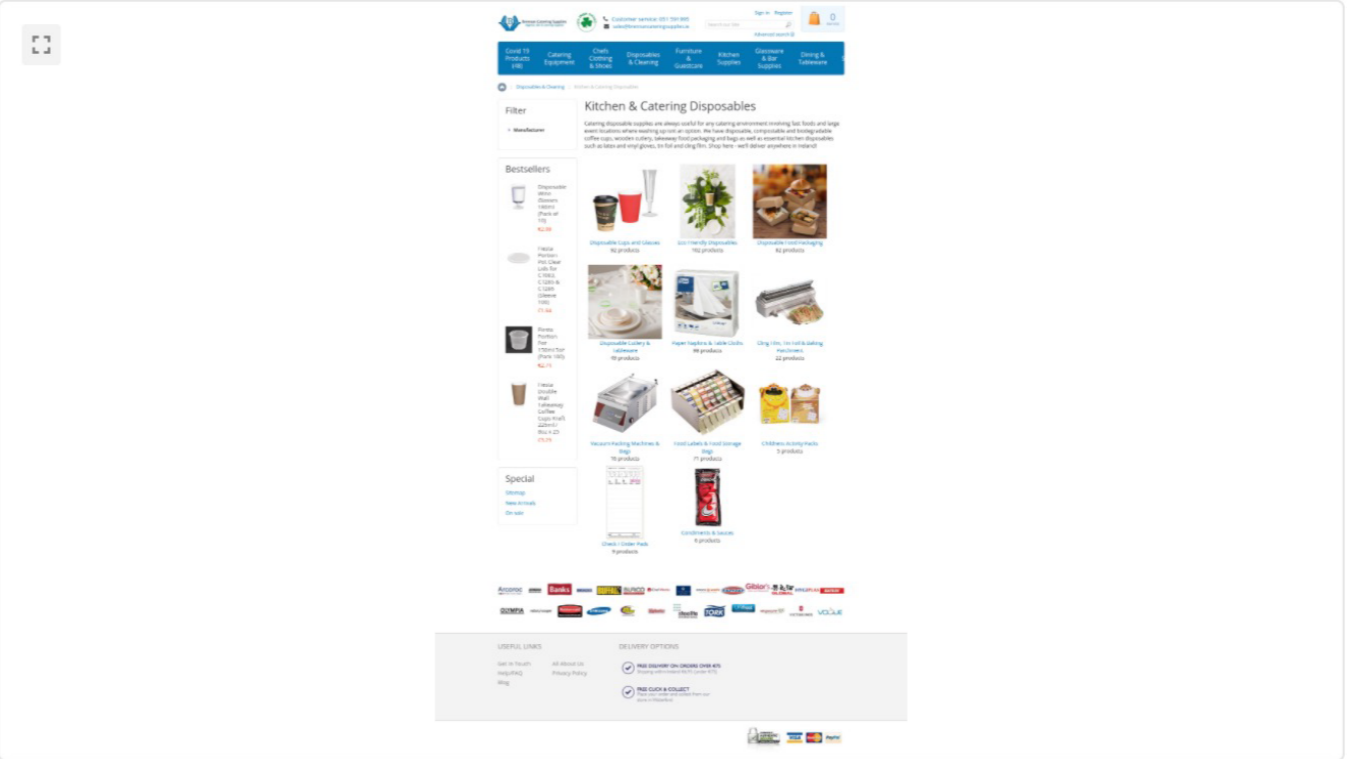
Query:
catering disposables

Region:
Dublin

A



B



Left link

Right link

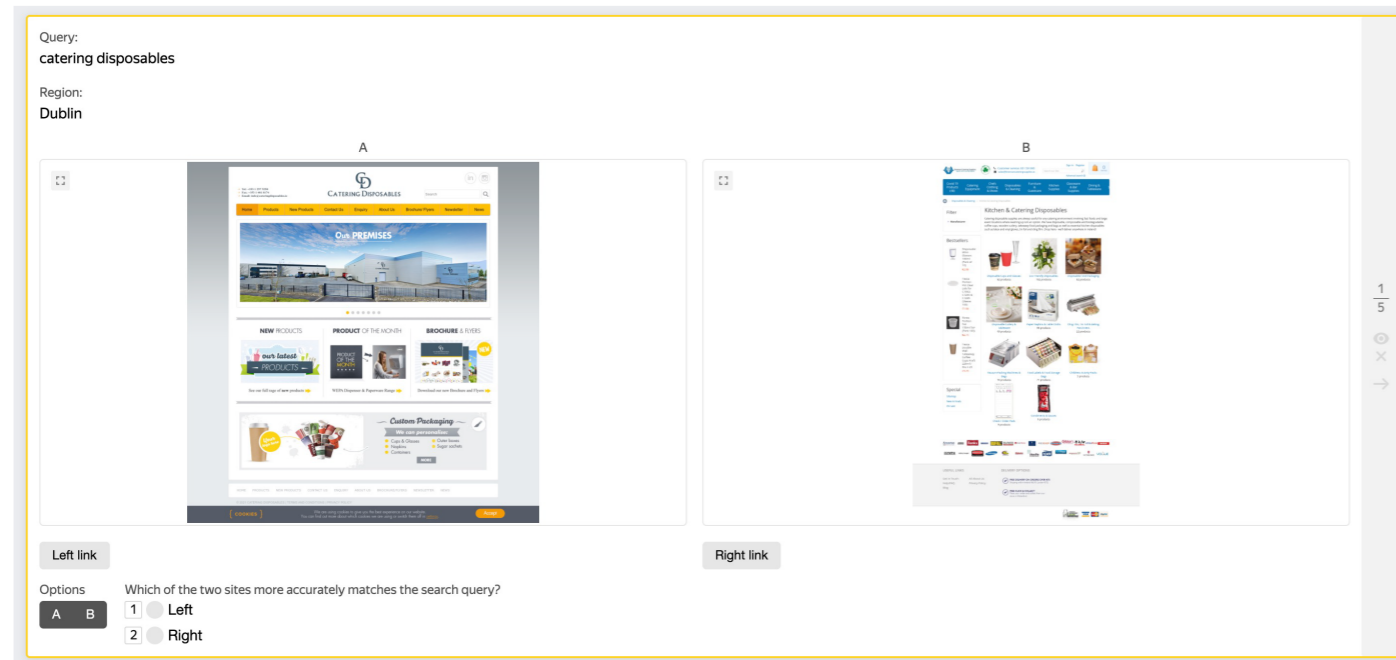
Options Which of the two sites more accurately matches the search query?

A B

1 Left

2 Right

Project:



Task

Which of the two sites more accurately matches the search query?

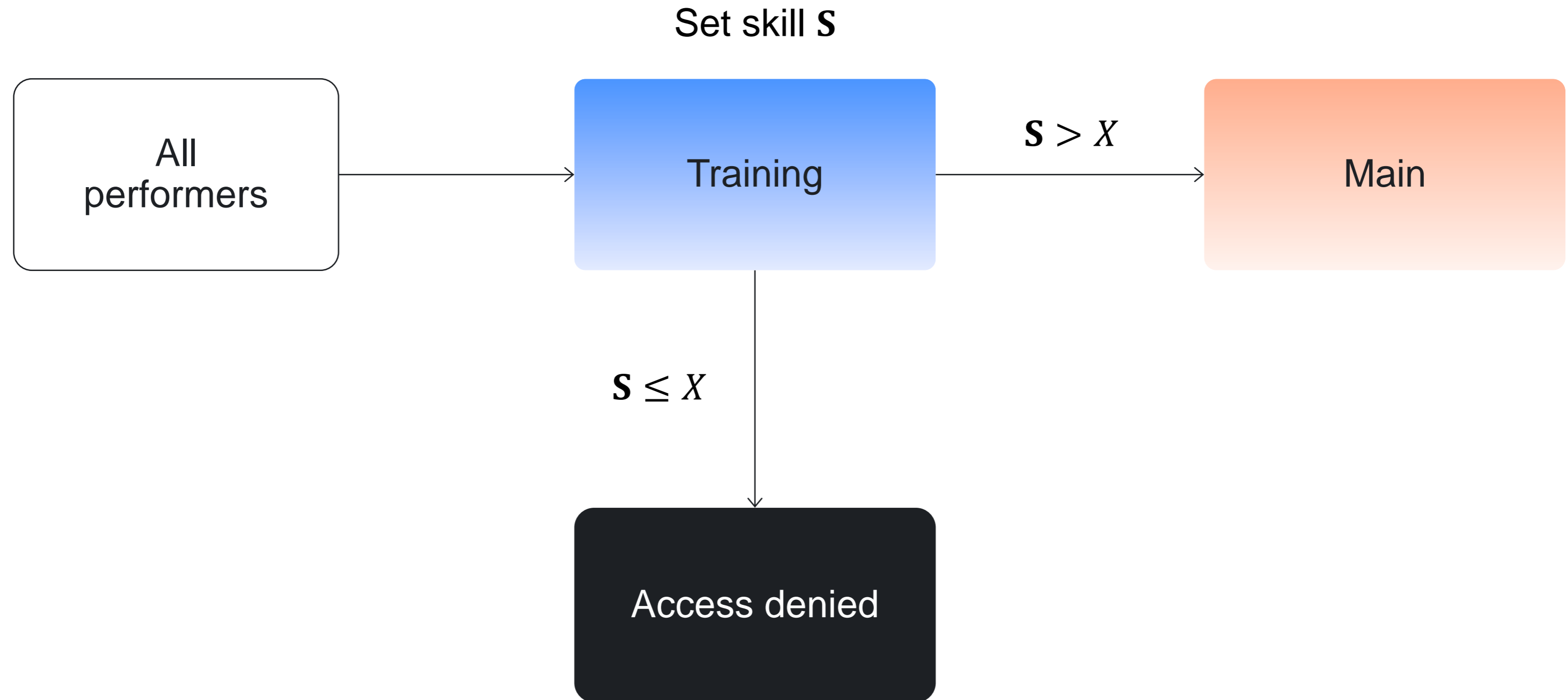
Key setting

- ▶ Type: Side-by-Side
- ▶ Training: yes
- ▶ Quality control: golden set
- ▶ Overlap: 6 or 8
- ▶ Pay: \$0.03 per 1 page with 5 tasks

Why?

Need to understand which URL works best for the query

Performer training



Toloka demo

5 min

What you need for the practice session

We are starting the practice session

We give you links to:

- ▶ A step-by-step instruction to configure and run our crowd project
- ▶ A dataset with queries and URLs
- ▶ Login+Password to sign in Toloka as a requester

Sign in Toloka as a requester

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

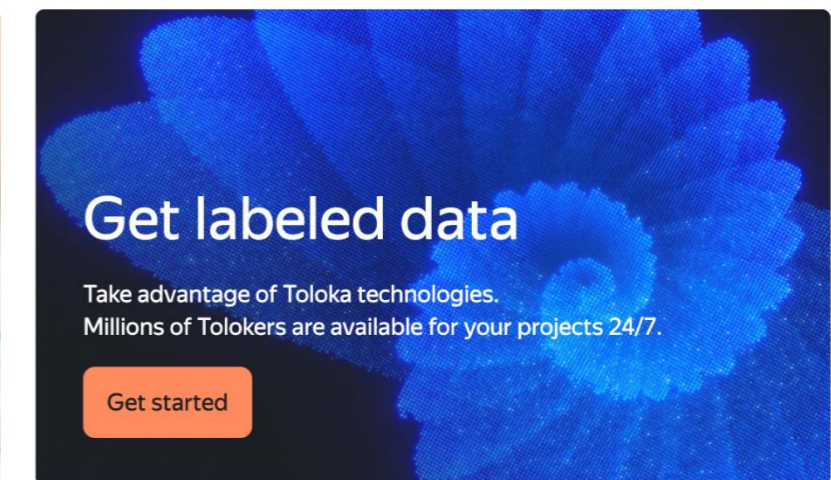
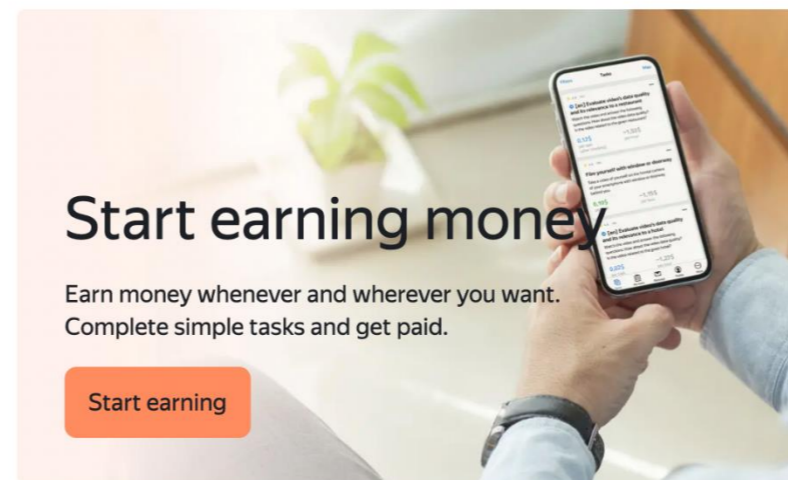


Sign in Toloka as a requester

3. Click on
“Get started”

4. Use received
Login+Password
to sign in

I want to...



Toloka account

Login

Password

In this part we create a project: we have 90 min

- ▶ You configure projects
- ▶ You set up training and main pool
- ▶ You run them on real performers
- ▶ We will run results aggregation together in Part V
- ▶ You have:
 - Instruction with all details
 - Dataset with queries and URLs

Thank you!

Questions?

Daria Baidakova

Sr. Manager, Customer Success & Education



dbaidakova@yandex-team.com



<https://research.yandex.com/tutorials/crowd/www-2021>