TABLE OF CONTENTS

PROJECT OVERVIEW

- **Objective:** Build a recommender system for Toloka, a gig working platform

- **Method:** Content-based approach, using NLP and algorithms to match job titles with relevant topics
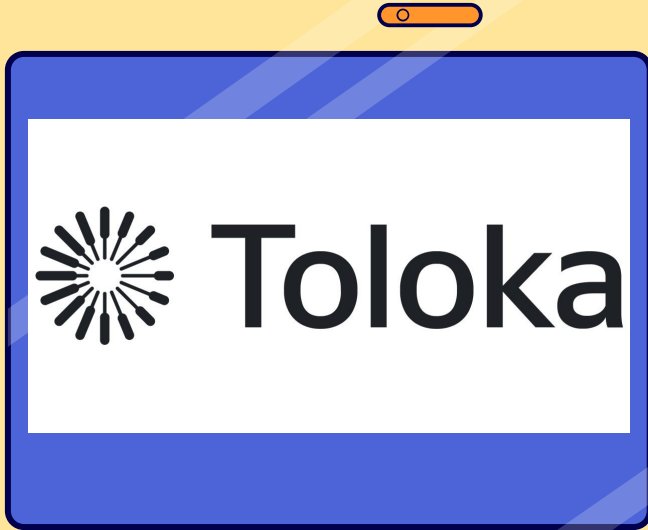
INITIAL RESULTS

Query: Check the similarity between two products

```
                                title
Binary product relevance markup
            MMRepresentativeImageV2
            Microsoft Bing App test
Product Inventory Validation V1
```

The output is not satisfactory, and the issue for was most probably presence of different languages in our corpus.

Language Detection

Issue: Job titles in different languages can be a problem for the algorithm

Solution: Used the 'Langdetect' library to detect and sort job titles by language

```
[ ] df_new1['language'].unique()

    array(['en', 'ru', 'ja', 'fr', 'es', 'mk', 'bg', 'ar', 'tr', 'pt'],
          dtype=object)
```

RESULTS

```
⤷  Query: Check the similarity between two products
    ----------------------------
    1.0   | Check the similarity between two products
    0.689 | Сравнение двух товаров в ритейле
    0.65  | How are the two products related? (v2)
    0.62  | Product Search Relevance Evaluation
    0.511 | Check if two audios has similar style
```

Russian - detected	⇄	English
сравнение двух товаров в ритейле	✕	comparison of two products in retail

Using the similarities between the titles only.

RESULTS

```
⤷  Query: Check the similarity between two products
   -----------------------------
   0.641 | Повторяющиеся товары
   0.635 | Check the similarity between two products
   0.625 | How are the two products related? (v2)
   0.577 | Product Image Validation V1
   0.564 | Какой текст лучше?
```

Russian - detected ▾	⇄	English ▾
Повторяющиеся товары ✕		Recurring Items

Using the similarities between the titles and descriptions, both.

Task Recommendations for Different Users

Addition: Task recommendations for different users based on their task history

Input: All tasks performed by each user

Output: Top 5 most relevant job titles for each user based on their task history

```
---------------------------
User ID: 1709755371.0
---------------------------
0.594 | Check the similarity between two products
0.575 | [Toloka] Test de compréhension de la langue française
0.564 | Яндекс Маркет. Редиректы. Категория. Точность
0.56 | Validation of English Ad Description
0.555 | Экзамен РЖЯ
---------------------------
User ID: 1718146027.0
---------------------------
0.727 | Earning money with Toloka
0.632 | Choosing the right tasks in Toloka
0.61 | Types of tasks in Toloka
0.559 | [Toloka] 日本語能力テスト
0.559 | [Toloka] Prueba de comprensión de español
---------------------------
User ID: 1718416044.0
---------------------------
0.799 | Choose how good the answer to the music query - Task in arabic
0.621 | Selection of performers for bounding box check tasks
0.61 | Selection of performers for text classification tasks
0.601 | Selection of performers for text and webpage classification tasks
0.59 | Selection of performers for text and image classification tasks
---------------------------
User ID: 1675921760.0
---------------------------
0.465 | ¿La captura de pantalla del cupón coincide con la información proporcionada del cupón? [MX-ES: Does coupon screenshot match the provi
0.461 | ¿La captura de pantalla del cupón coincide con la información proporcionada del cupón? [ES: Does coupon screenshot match the provided
0.447 | Does coupon screenshot match the provided coupon's information? (USA)
0.446 | Tours and Activities Ads Relevance - Portuguese
0.444 |   La capture d'écran du coupon correspond-elle aux informations du coupon fourni ?
```

id	titles	EstimateType	Budget	jobHours	Duration	length_AdPosted	wages	application_cost
"159758 0572455 288832"	Looking for a website developer or builder to help finish my website</a	budget	300			4 hours ago	Fixed-price	2 Connects
"159757 9526850 125824"	Voice over freelancer needed for part-time services</a	duration		30 hrs/week	Less than 1 month, Less than	4 hours ago	Hourly	1 Connect
"159757 9437450 346496"	Small MAC file management project</a	duration		30 hrs/week	Less than 1 month, Less than	4 hours ago	Hourly: $31.00-$50.00	1 Connect
"159757 7870564 343808"	Cold Email Copywriter for Sales Development Rep Prospecting</a	duration		30 hrs/week	Less than 1 month, Less than	4 hours ago	Hourly: $50.00-$86.00	2 Connects
"159757 7397461 557248"	Virtual assistance needed for customer support.</a	duration		30 hrs/week	Less than 1 month, Less than	4 hours ago	Hourly	1 Connect

Data from Upwork

FUTURE WORK

Goal: Incorporate collaborative filtering into the recommender system

Method: Collaborative filtering approach to suggest tasks based on user profiles and preferences

THANK YOU!