Annotation Guidelines for Reviewing Answer Matching for Lithuanian

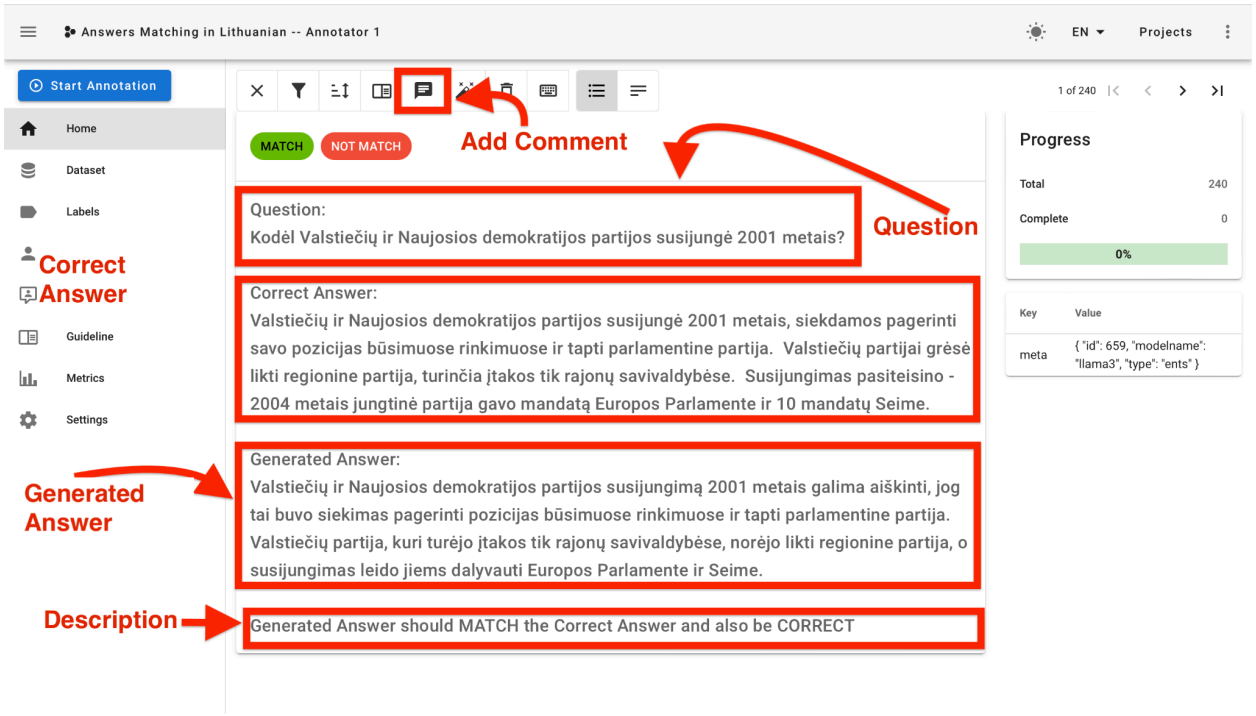
The goal of this annotation task is to review and label triplets of *questions*, *correct answers*, and *generated answers* according to their alignment and correctness.

Motivation: this dataset will help to evaluate AI responses on open-ended questions in Lithuanian and could be utilized for LIthuanian Automated Short Answer Grading (ASAG) for the education industry.

You will be presented with (see the image below):

1. A **Question** (in Lithuanian).
2. A **Correct Answer** (the expected answer to the question, used as a reference).
3. A **Generated Answer** (produced by an AI system).
4. A **Description**, specifying one of two scenarios:
    - **Scenario 1:** “Generated Answer should MATCH the Correct Answer and also be CORRECT.”
    - **Scenario 2:** “Generated Answer should NOT MATCH the Correct Answer and also be INCORRECT.”

Based on the input, you need to select one of the labels: **MATCH** or **NOT MATCH**. Also, you need to **write a Comment** in English on why you selected the **NOT MATCH** label.



Your task is to evaluate:

- If the **Generated Answer** aligns with the **Correct Answer** (in meaning, accuracy, and relevance).
- Whether the **Generated Answer** fulfills the criteria specified in the **Description**.

Based on your evaluation, assign one of the following labels:

- **MATCH:** The **Generated Answer** aligns with the **Correct Answer** and follows the scenario in the description.
- **NOT MATCH:** The **Generated Answer** does not align with the **Correct Answer** or does not follow the scenario in the description. ***If you assign this label - please, add comment on why you did it in English.***

Possible Options:

1. **Generated Answer** aligns with the **Correct Answer** (what alignment means will be provided further with examples) and **Scenario 1.**

   In this case: Assign the label **MATCH.**

2. **Generated Answer** *DOES NOT* align with the **Correct Answer** (what alignment means will be provided further with examples) and **Scenario 1.**

   In this case: Assign the label **NOT MATCH** and write a **Comment** on why**.**

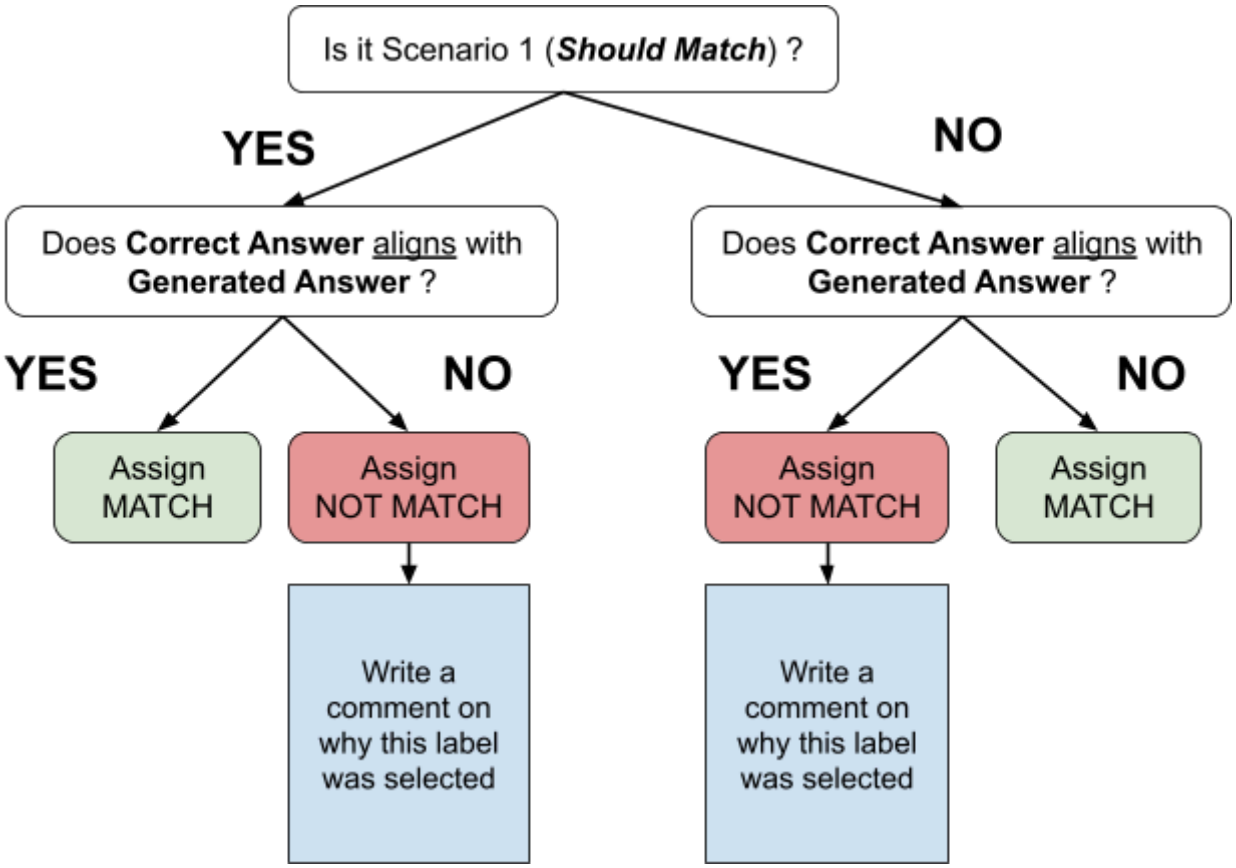
3. **Generated Answer** align with the **Correct Answer** (what alignment means will be provided further with examples) and **Scenario 2.**

   In this case: Assign the label **NOT MATCH** and write a **Comment** on why**..**

4. **Generated Answer** *DOES NOT* align with the **Correct Answer** (what alignment means will be provided further with examples) and **Scenario 2.**

   In this case: Assign the label **MATCH.**

For simplicity, you can use the following decision tree:



General Rules on the alignment of the Correct Answer and Generated Answer:

1. **Evaluate Meaning, Not Exact Words:** Focus on whether the meaning of the **Generated Answer** matches the **Correct Answer**. The wording does not need to be identical, but the core information should align.

2. **Typos**: typos are acceptable in both **Generated Answer** and **Correct Answer.** If the typo is changing the meaning significantly or makes the text unreadable, assign **NOT MATCH.** Otherwise, if the typos are presented, but the core information is matching - the answers are aligned.

3. **Relevance Matters:** The **Generated Answer** must address the question. Irrelevant information or incomplete answers should be labeled as **NOT MATCH**. However, the **Generated Answer** could contain more information than **Correct Answer,** but **Generated Answer** should contain all the relevant bits as well**.**

4. **Additional information.** Sometimes**, Correct Answer** could include relevant bits of information, but which could be omitted when answering the question. If those details are missing from the **Generated Answer,** but **Generated Answer** answers the question and does not contradict the **Correct Answer** - they are **aligned.**

5. **Scenario:** Always consider the provided **Description**. An answer may match the correct answer but still be labeled **NOT MATCH** if it contradicts the scenario's requirements.

Examples

Example 1

Question:

Kas yra Lietuvos sostinė?

Correct Answer:

Vilnius.

Generated Answer:

Lietuvos sostinė – Vilnius!

Generated Answer should MATCH the Correct Answer and also be CORRECT.

Evaluation:
The **Generated Answer** matches the **Correct Answer** and satisfies the scenario.
Label: MATCH.

Example 2

Question:

Kada ir kieno buvo sukurta pirmoji lietuviška abėcėlė?

Correct Answer:

Pirmąją lietuvišką abėcėlę 1547 m. parengė Martynas Mažvydas. Joje buvo 23 gotiškojo šrifto raidės, tačiau dar nebuvo raidžių č, š, ž, ą, ę, ė, į, ų, ū.

Generated Answer:

Pirmoji lietuviška abėcėlė buvo sukurti 1547 metais Martyno Mažvydo iniciatyva.

Generated Answer should MATCH the Correct Answer and also be CORRECT.

Evaluation:
The **Generated Answer** matches the **Correct Answer** and satisfies the scenario. Note

that the **Correct Answer** has more information towards the question, but the **Generated Answer** fully answers it without those additional facts.
Label: MATCH.

Example 3

Question:

Kas yra Lietuvos sostinė?

Correct Answer:

Vilnius.

Generated Answer:

Lietuvos sostinė yra Vilnius ir Kaunas

Generated Answer should MATCH the Correct Answer and also be CORRECT.

Evaluation:
The **Generated Answer** *does match*, as it provides information that contradicts the **Correct Answer**. Therefore, it does not satisfy the scenario.
Label: NOT MATCH.

Possible Comment: Vilnius was in the generated answer, but the model also said that Kaunas is a capital as well, which contradicts the answer.

Example 4

Question:

Kas yra Lietuvos sostinė?

Correct Answer:

Vilnius.

Generated Answer:

Lietuvos sostinė yra Vilnius ir Kaunas

Generated Answer should NOT MATCH the Correct Answer and also be INCORRECT.

Evaluation:
The **Generated Answer** *does match*, as it provides information that contradicts the **Correct Answer**, therefore it satisfies the description.
Label: MATCH.

Example 5

Question:

Kas yra Lietuvos sostinė?

Correct Answer:

Vilnius.

Generated Answer:

Lietuvos sostinė yra Vilnius :)

Generated Answer should NOT MATCH the Correct Answer and also be INCORRECT.

Evaluation:
The **Generated Answer** *match*, as it provides information that contradicts the **Correct Answer**, therefore it satisfies the description.
Label: NOT MATCH.

Possible Comment: Generated answer and correct answer are the same.

Example 6

Question:

Kas yra Lietuvos sostinė?

Correct Answer:

Vilnius.

Generated Answer:

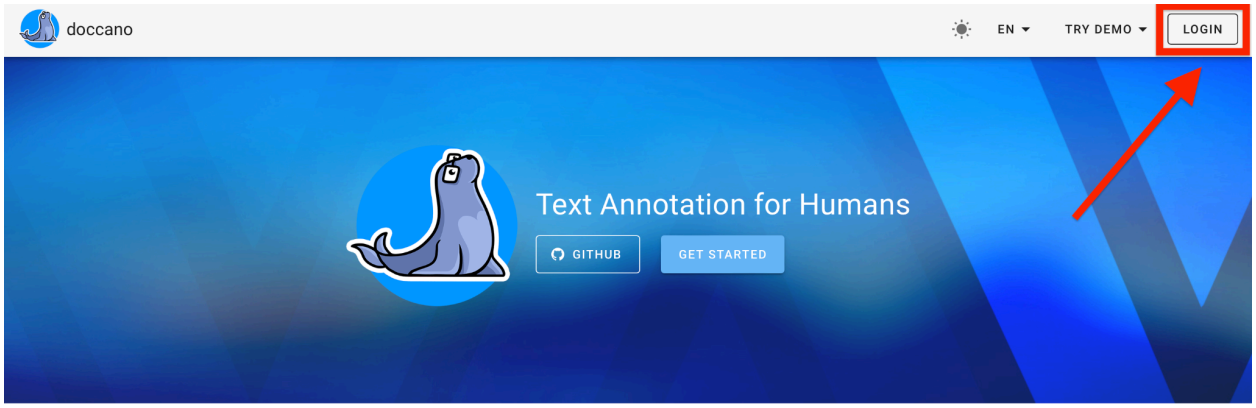
Lietuva turi tiek daug puikių miestų: Vilnius, Kaunas, Klaipėda ir panašiai, bet pirmasis yra sostinė.

Generated Answer should MATCH the Correct Answer and also be CORRECT.

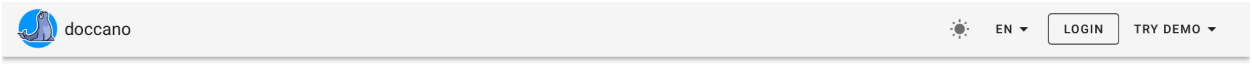
Evaluation:
The **Generated Answer** matches the **Correct Answer** and satisfies the scenario. Note that it provides more information than the original answer, but it does not contradict it.
Label: MATCH.

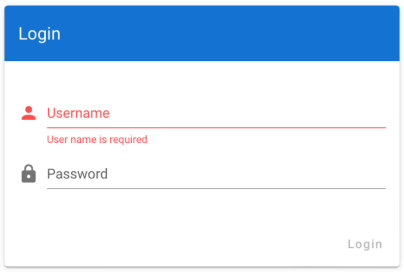
System login and overview

1. Go to the: http://35.223.56.105:80/ and click Login

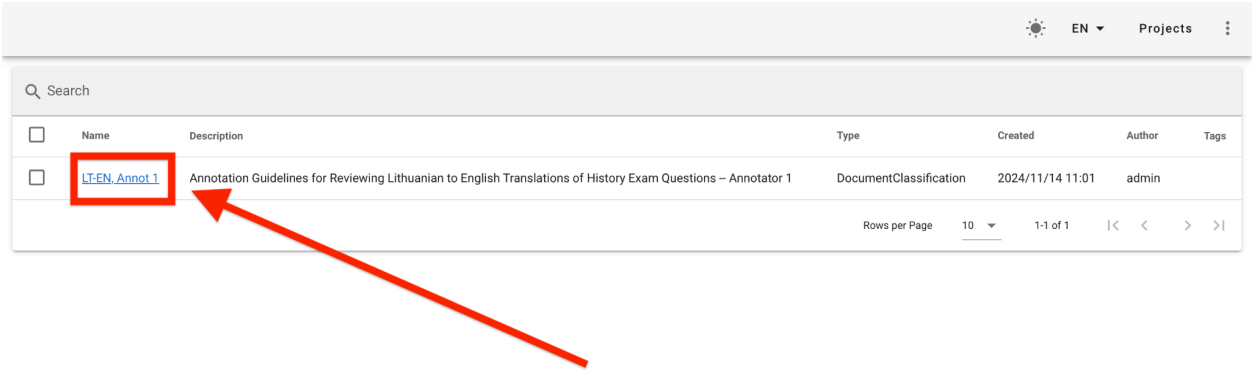


2. You will be asked to put your username and password (it should be provided beforehand). Put them and click Login.

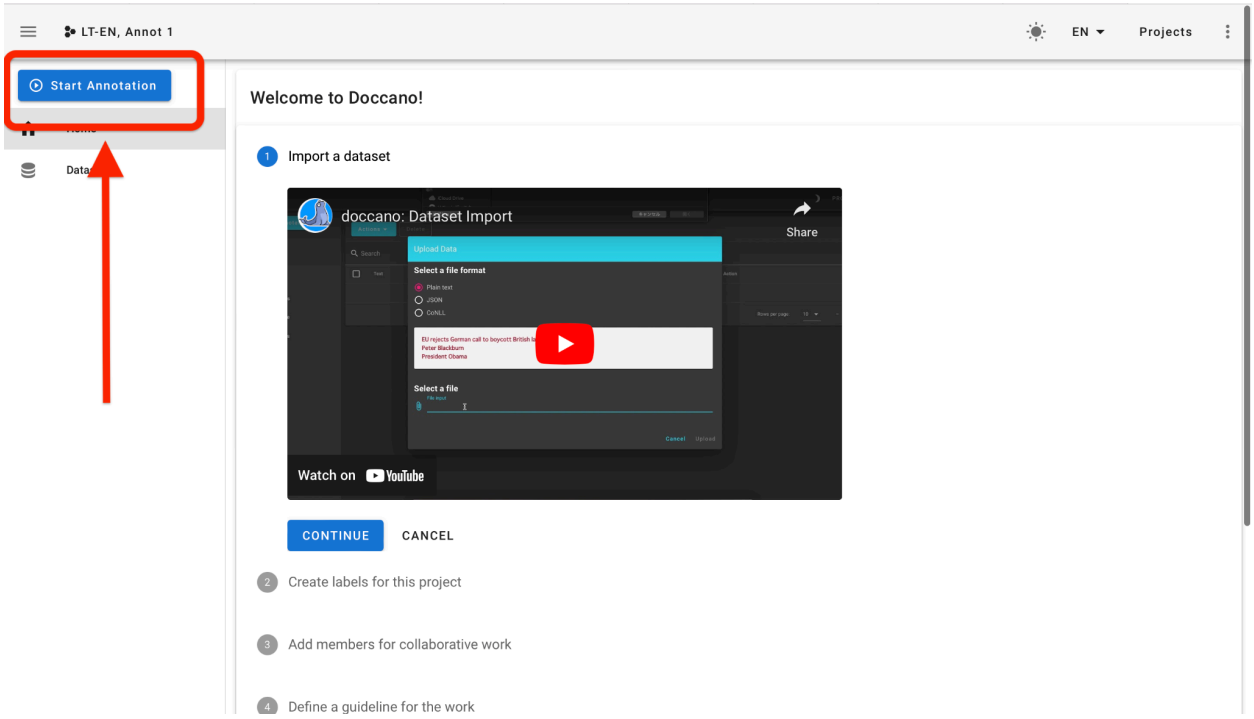




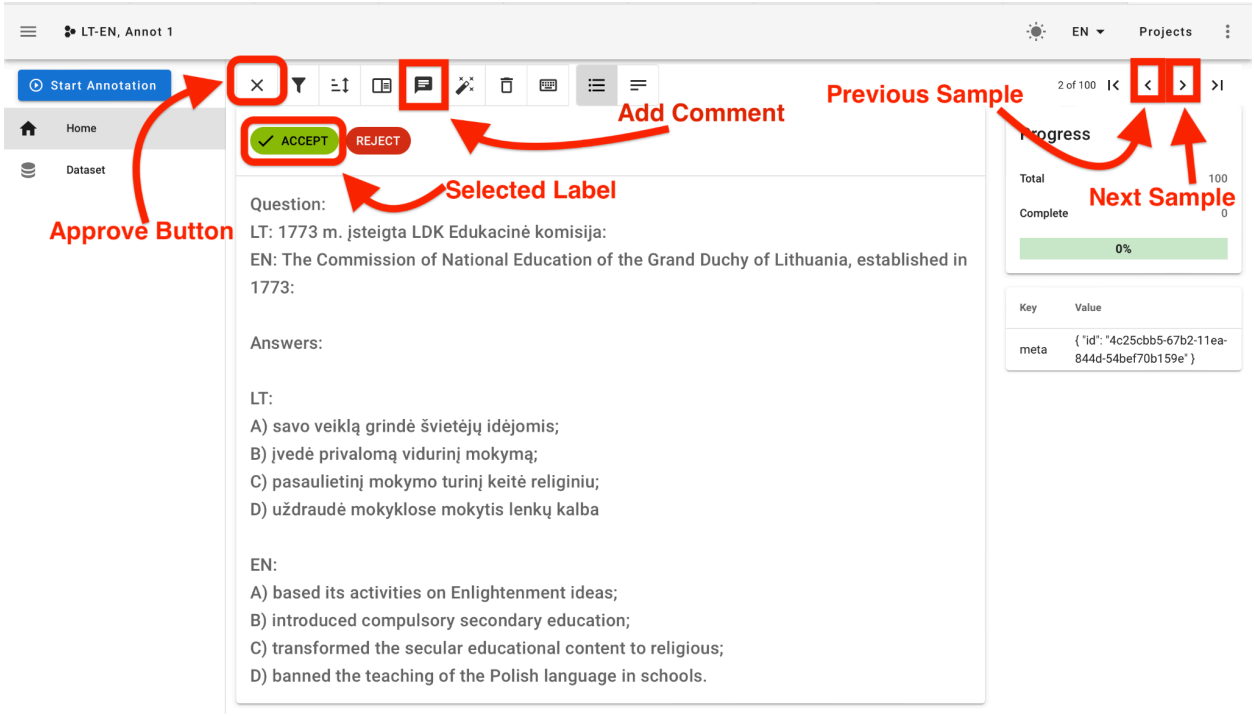
3. Once you logged in, you will be redirected to all the projects you are a part of as an annotator. **Select the one that has a name “Answers Matching in Lithuanian”.**



4. You will be redirected to the annotation page. Click “Start Annotation”.



5. Finally, it will open the last sample you stopped annotating last time. To annotate a sample, read the text and **click on the appropriate label (see in the example below)**. Once the label is chosen, there will be a mark near it (see image below). To go to the next sample, click “>” on the top right side. To go to the previous sample, click “<”. Once you click on the label and go to the next sample, your annotation is saved.
   Also, optionally, you can click the “Approve” button, which will update the percentage counter of how much was annotated. **See image below.**


LT-EN, Annot 1
EN
Projects
Start Annotation
Home
Dataset
2 of 100
ACCEPT
REJECT
Add Comment
Previous Sample
Next Sample
Approve Button
Selected Label
Question:
LT: 1773 m. įsteigta LDK Edukacinė komisija:
EN: The Commission of National Education of the Grand Duchy of Lithuania, established in 1773:
Answers:
LT:
A) savo veiklą grindė švietėjų idėjomis;
B) įvedė privalomą vidurinį mokymą;
C) pasaulietinį mokymo turinį keitė religiniu;
D) uždraudė mokyklose mokytis lenkų kalba
EN:
A) based its activities on Enlightenment ideas;
B) introduced compulsory secondary education;
C) transformed the secular educational content to religious;
D) banned the teaching of the Polish language in schools.
Total
100
Complete
0
0%
Key
Value
meta
{ "id": "4c25cbb5-67b2-11ea-844d-54bef70b159e" }