Knowledge Engineering Report

Edoardo Saturno, Riccardo Marvasi
Master's Degree in Artificial Intelligence, University of Bologna
riccardo.marvasi@studio.unibo.it edoardo.saturno@studio.unibo.it

1 Introduction

We utilized an existing tool for knowledge engineering called the **OntoChat** model. This tool is publicly available on GitHub [1] and has a shared platform where it can be used by anyone remotely. We worked on it locally in order to be able to change the original files and modify the tool.
OntoChat is particularly useful for various tasks related to **ontology construction, analysis, and improvement**. It integrates with the OpenAI API, utilizing a public key for access to one of their models.

Our project aim was to ensure that ontochat tool had the following features:

- Evaluation of an ontology with respect to a series of Competency Question
- Automatic creation of ontology drafts based on given Competency Questions

2 System Description

2.1 Ontology Story creation

This feature enables users to create detailed and structured stories that describe how an ontology will be used in practical scenarios. The process involves guiding users through various stages:

- **Persona Definition**: Users create personas, which are fictional characters that represent typical users of the ontology. These personas are detailed with information such as name, occupation, age, skills, and interests.
- **Goals and Scenarios**: Users then define clear objectives and situations where the ontology will be applied, taking into account how the persona interacts with the system.
- **Iterative Refinement**: The system refines these stories through prompts and feedback loops, asking the user questions and building a comprehensive narrative. As needed, the LLM will continue to elicit details from the user until a sufficiently rich story is developed

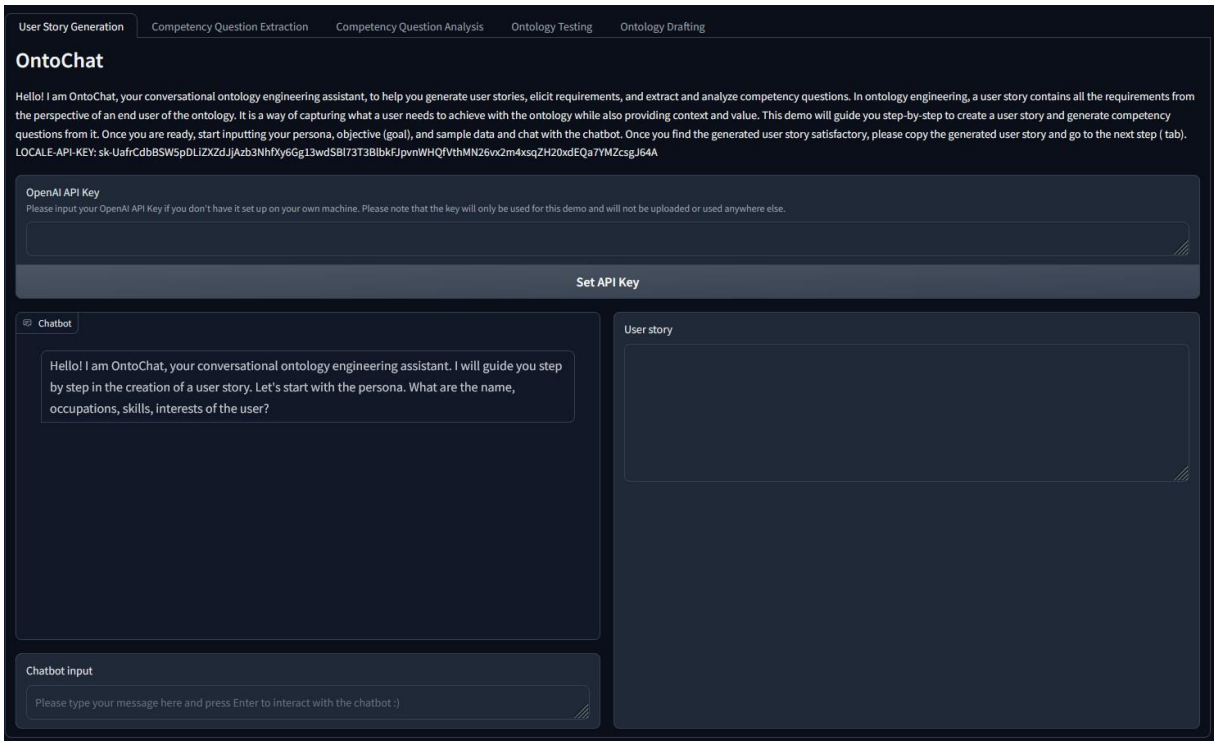

Figure 1

2.2 Ontology QA generation

Competency Questions (CQs) are essential for defining the requirements and constraints of an ontology. OntoChat helps extract CQs directly from user stories. This is done in several steps:

- **Initial CQ Extraction**: After the user provides a story, OntoChat suggests an initial set of CQs that address the key goals and constraints of the ontology.
- **Refinement of CQs**: It then refines these questions by breaking down complex or nested CQs into simpler, atomic ones. For example, if a CQ is too broad, like “What styles and genres are associated with the song Penny Lane?”, it would split it into two separate CQs, such as “What genres are associated with the song?” and “What styles are associated with the song?”.
- **Abstraction**: Named entities like “Penny Lane” are replaced with abstract entities, making the questions more generalized and easier to reuse in different contexts

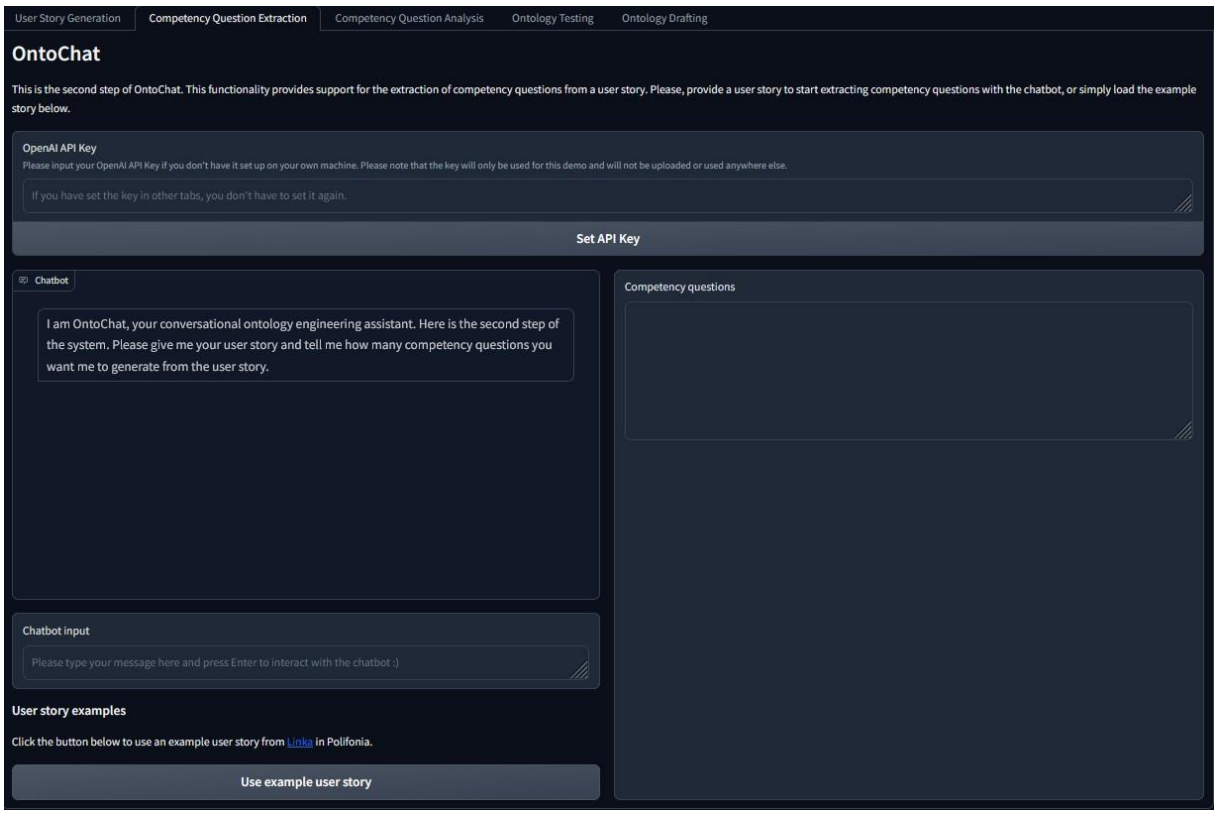

Figure 2

2.3 Ontology Analysis

This feature helps users refine the extracted CQs by removing redundancies and organizing them into meaningful clusters:

- **Paraphrase Identification**: The system automatically detects and eliminates redundant questions, ensuring that only unique and meaningful CQs remain.
- **Clustering**: OntoChat clusters similar CQs, grouping them by thematic focus or intent. This helps ontology engineers better understand the requirements and makes it easier to assign tasks based on specific clusters. For instance, one team may handle a group of CQs related to taxonomy, while another focuses on data interoperability

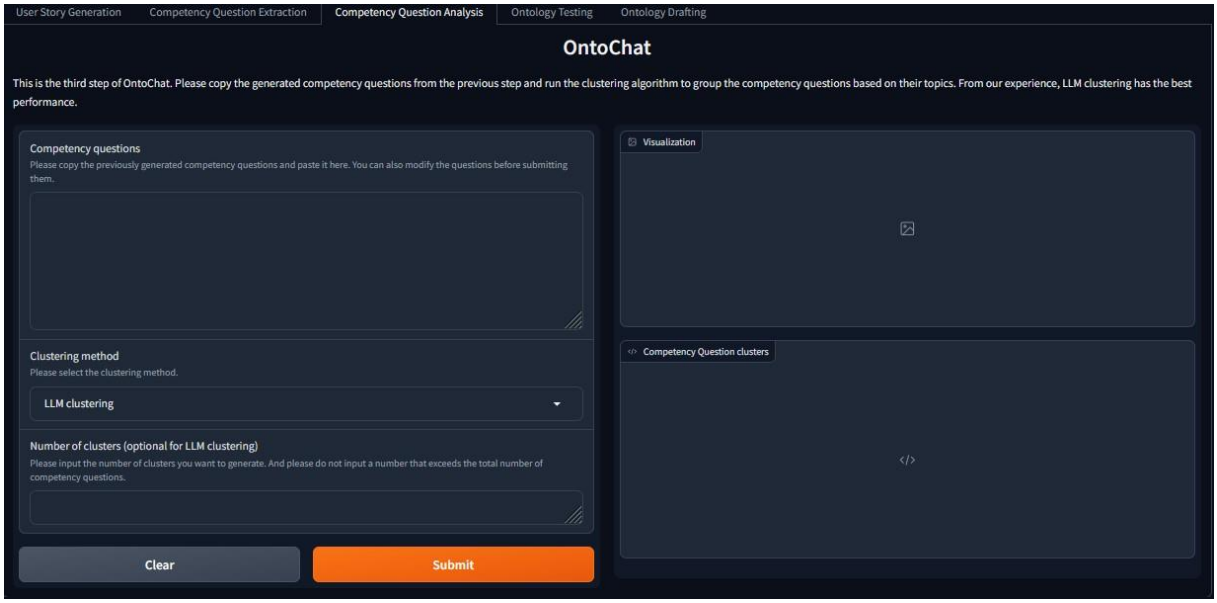

Figure 3

2.4 Ontology Testing

This feature focuses on verifying whether the generated ontology meets the specified requirements:

- **Competency Question Verification**: CQs are formalized into queries (such as SPARQL queries) to check if the ontology provides correct answers. This helps ensure that the ontology covers the needed competencies.
- **Inference Testing**: This verifies if the ontology can logically infer new knowledge based on existing data.
- **Error Provocation**: The system attempts to identify weak points in the ontology by intentionally creating scenarios where the ontology might fail. This helps engineers fix potential issues early on.

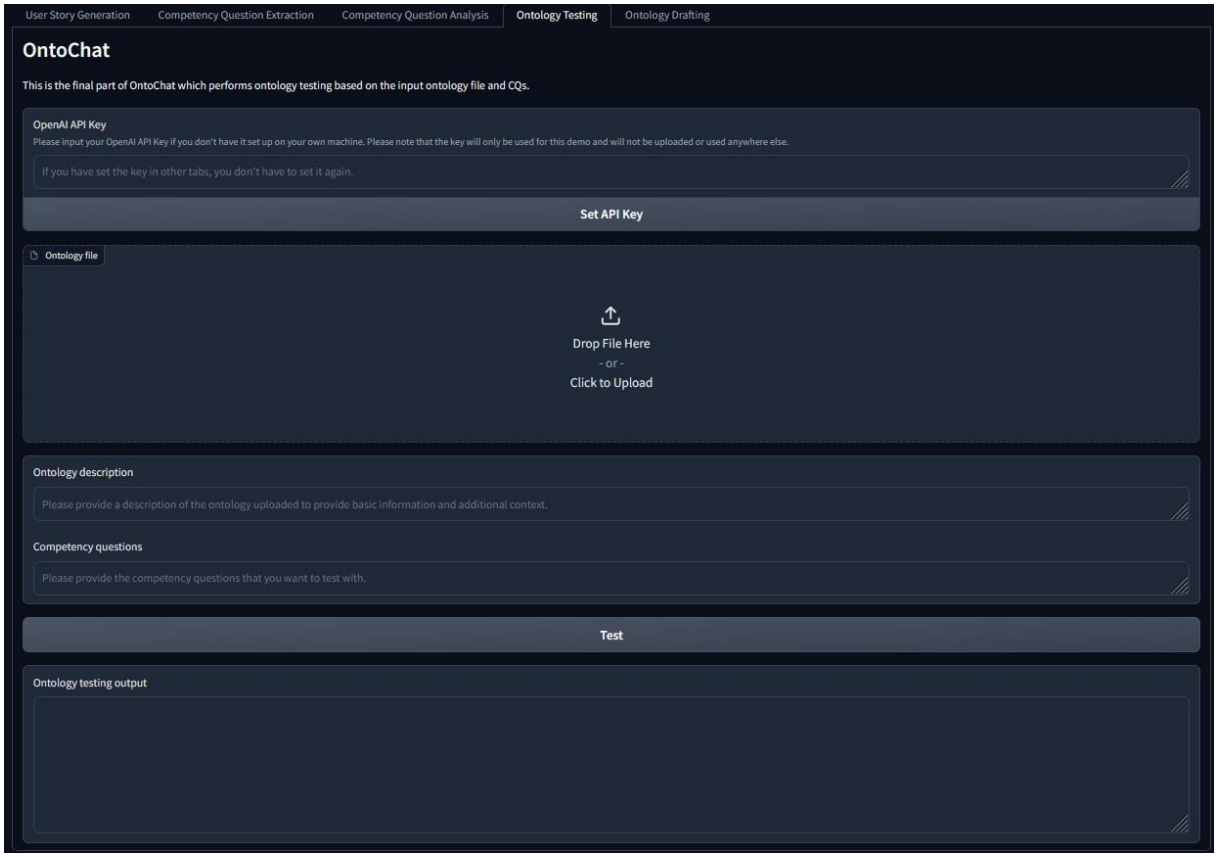

Figure 4

2.5 Ontology RDF generation

In this phase, a user story is input into the model, which processes it to generate competency questions. The questions are then utilized to create two types of files:

- **RDF files**, which are machine-readable and can be used in Step 4 for testing against the competency questions.
- **Turtle files**, which are human-readable representations of the ontology.

Both file types represent draft ontologies, which can later be refined and enhanced by the user and used for testing the ontology.

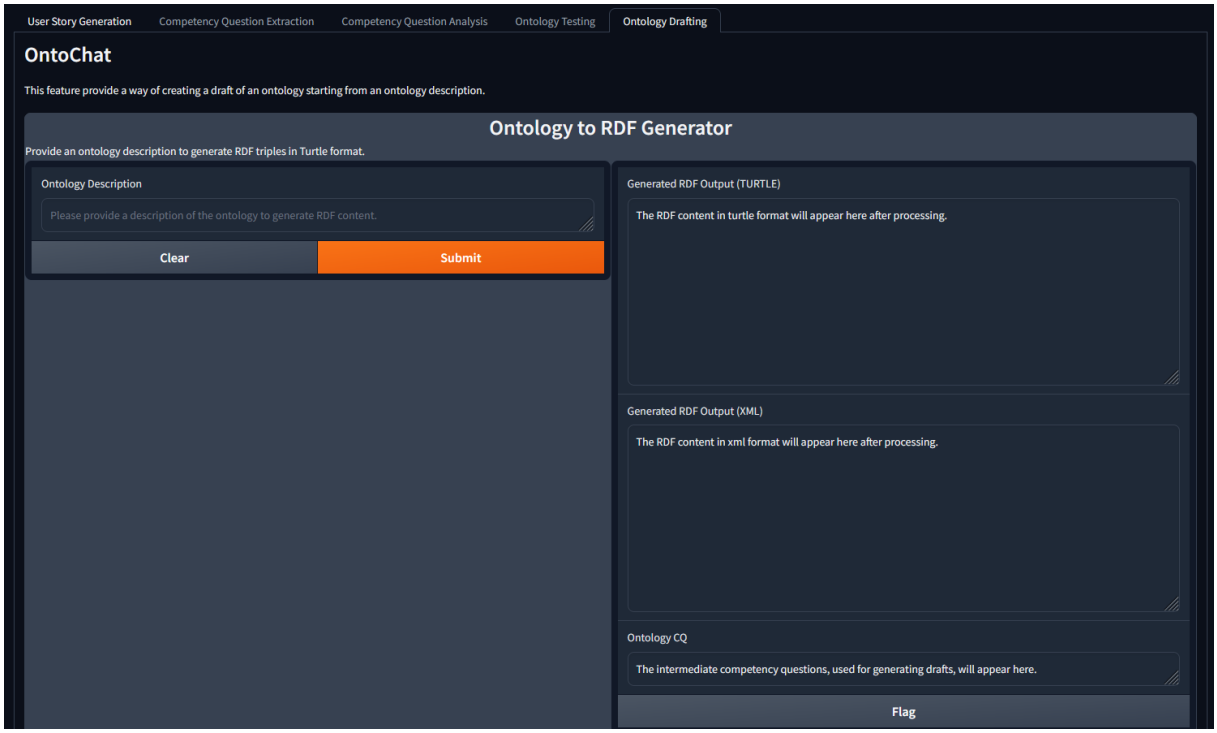

Figure 5

3 Data

We did not utilize any external datasets, as we relied on the OpenAI API, which provides a pre-trained and fine-tuned model capable of functioning as a large language model (LLM). However, we employed intermediate text files for testing purposes, which can be modified to address various scenarios. Specifically, we used three different test stories invented by us. These files allow us to simulate the experience of endpoint users and evaluate the tool's results.

The first story takes place in a fantasy setting, it is described as follows:

Edoardo is one of the greatest warlords to ever walk the Badlands. He wants to organize an arena event called "BloodBath," where the strongest orcs of the Badlands will test their skills against one another. Each fighter is either a "Black Orc" or an "Orc Boyz" and each carries a specific weapon. The weapon types include "Great Axe," "Choppa," or "Big Rusted Sword". Additionally, some fighters may randomly take on the role of a "Black Orc Big Boss" if they are Black Orcs or an "Orc Big Boss" if they are Orc Boyz. An Orc Big Boss of any type is bigger and stronger than a regular orc and is able to kill opponents in one hit.

The second story is related to motorcycles races:

The Piston Cup is an event organized by Riccardo that takes place every year in Oregon, starting from 2016. Two types of motorcycles participate in the competition: the white motorcycle and the red motorcycle. The white motorcycle has a very high top speed but has a low acceleration, while the red motorcycle has excellent acceleration but a low top speed. Each year, the track has a different shape; sometimes there are many curves, while other times it is entirely straight. It is boss Riccardo Marvasi who decides what shape the track will have from year to year. Another interesting aspect of the Piston Cup is that sometimes it rains heavily. When it rains, the red motorcycle is greatly advantaged, while when it is very hot and there is no rain, the white motorcycle has the advantage.

The third story is about a retail worker:

Giovanni works in retail and wants to earn more money, but he can only do so if he works for Company A. To work at Company A, he must have retail experience with another company. Currently, Giovanni works for Company B, so he has the necessary experience to make the move. However, Giovanni strongly dislikes Company C and does not want any involvement with it or its related companies. Companies can either be associated or not associated with Company C, which is known for hiring both experienced and inexperienced retail workers. Additionally, Giovanni has a friend who works for Company C.

We will conduct the following reasoning to demonstrate that the tests are coherent and sound with respect to the tool:

- First, we will show that by generating the competency questions directly from the user story using the **Competency_questions_extraction** feature, we obtain the same results as when generating them from the draft ontologies created via the ontology drafting feature.
- Secondly, we will demonstrate that the model correctly responds to three different groups of competency questions. The first group is generated using the QA extraction feature, the second group comes from the ontology drafting feature, and the third group is based on the same method as the second but includes randomly substituted questions that are unrelated to the ontology structure. We will input the RDF file generated from the **ontology_drafting** feature, along with the QA-generated questions and the context of the story.

4 Results

4.1 First Test

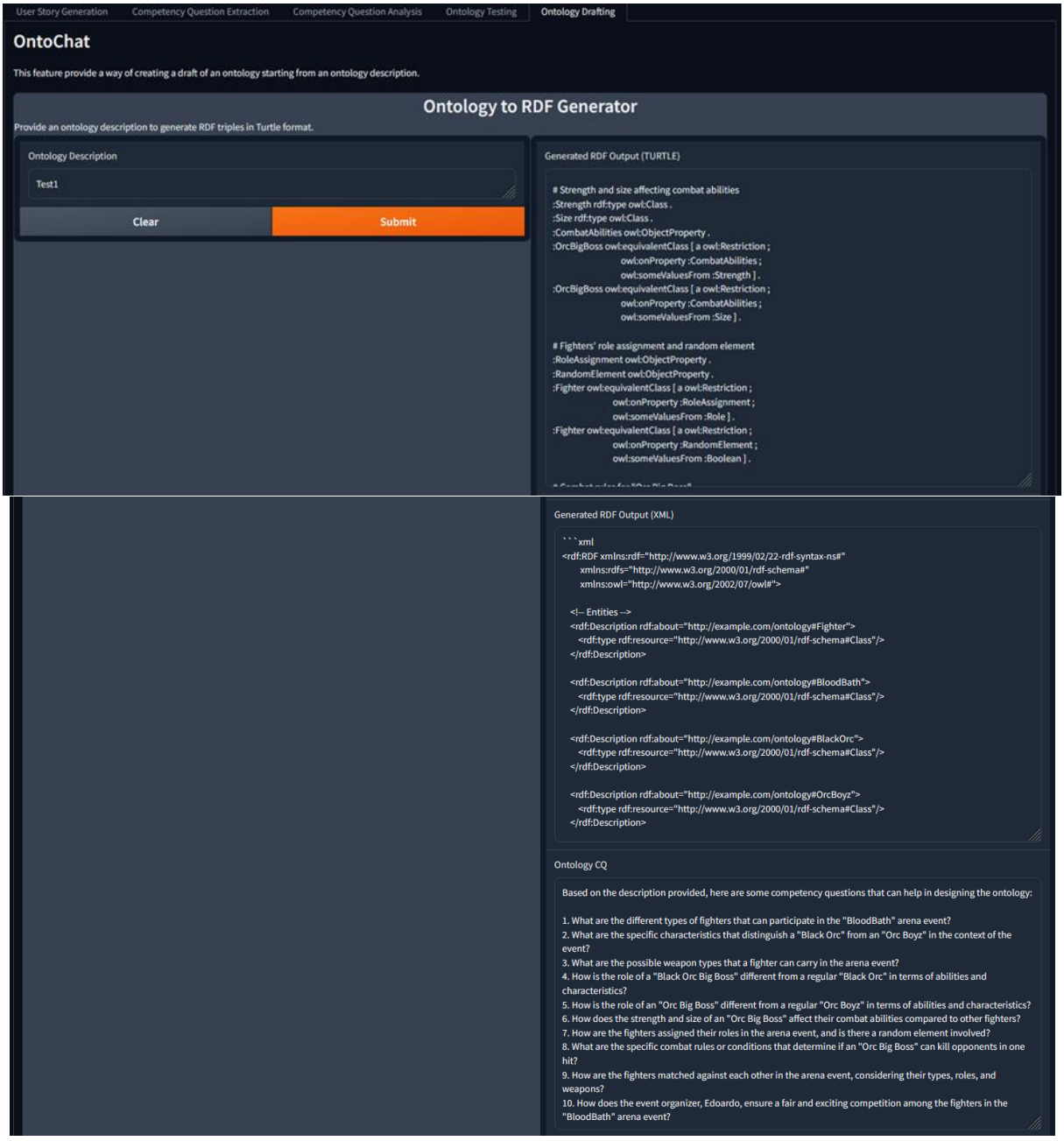

Figure 6

These are the CQ extracted from the **ontology_drafting** features:

1. *What are the different types of fighters that can participate in the "BloodBath" arena event?*
2. *What are the specific characteristics that distinguish a "Black Orc" from an "Orc Boyz" in the context of the event?*
3. *What are the possible weapon types that a fighter can carry in the arena event?*

4. *How is the role of a "Black Orc Big Boss" different from a regular "Black Orc" in terms of abilities and characteristics?*
5. *How is the role of an "Orc Big Boss" different from a regular "Orc Boyz" in terms of abilities and characteristics?*
6. *How does the strength and size of an "Orc Big Boss" affect their combat abilities compared to other fighters?*
7. *How are the fighters assigned their roles in the arena event, and is there a random element involved?*
8. *What are the specific combat rules or conditions that determine if an "Orc Big Boss" can kill opponents in one hit?*
9. *How are the fighters matched against each other in the arena event, considering their types, roles, and weapons?*
10. *How does the event organizer, Edoardo, ensure a fair and exciting competition among the fighters in the "BloodBath" arena event?*

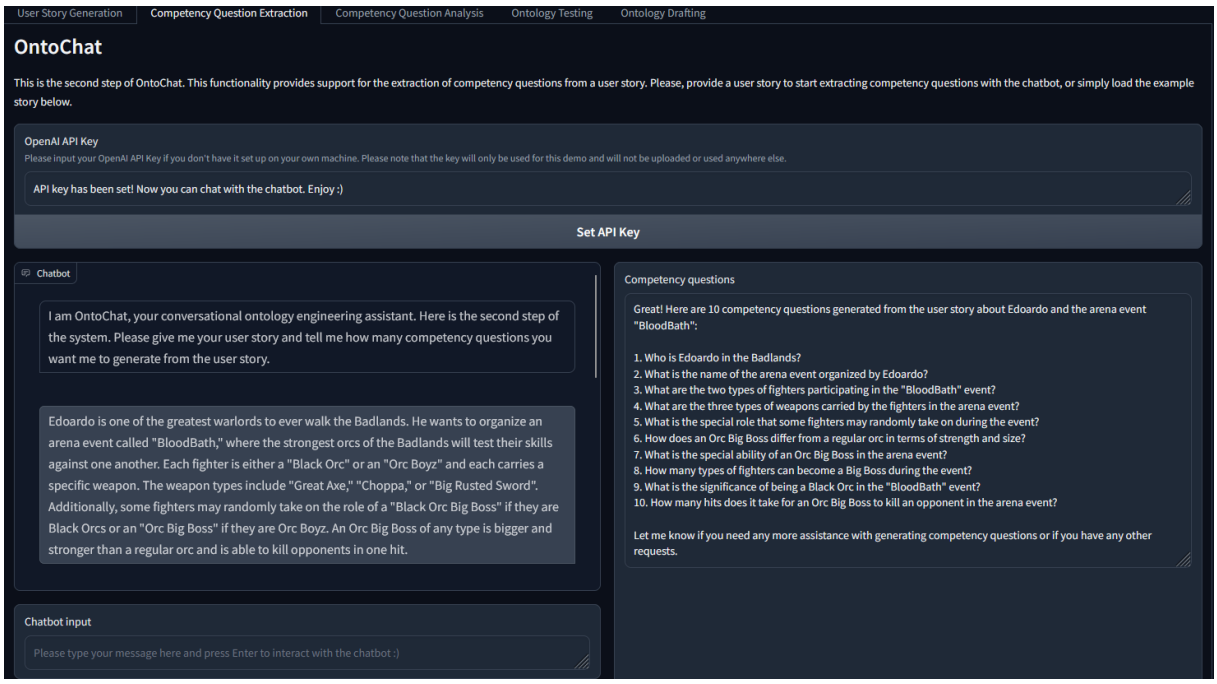

Figure 7

These are the CQ extracted from **Competency_questions_extraction** features, as seen in figure 7:

1. *Who is Edoardo in the Badlands?*
2. *What is the name of the arena event organized by Edoardo?*
3. *What are the two types of fighters participating in the "BloodBath" event?*
4. *What are the three types of weapons carried by the fighters in the arena event?*
5. *What is the special role that some fighters may randomly take on during the event?*
6. *How does an Orc Big Boss differ from a regular orc in terms of strength and size?*
7. *What is the special ability of an Orc Big Boss in the arena event?*
8. *How many types of fighters can become a Big Boss during the event?*

9. *What is the significance of being a Black Orc in the "BloodBath" event?*
10. *How many hits does it take for an Orc Big Boss to kill an opponent in the arena event?*

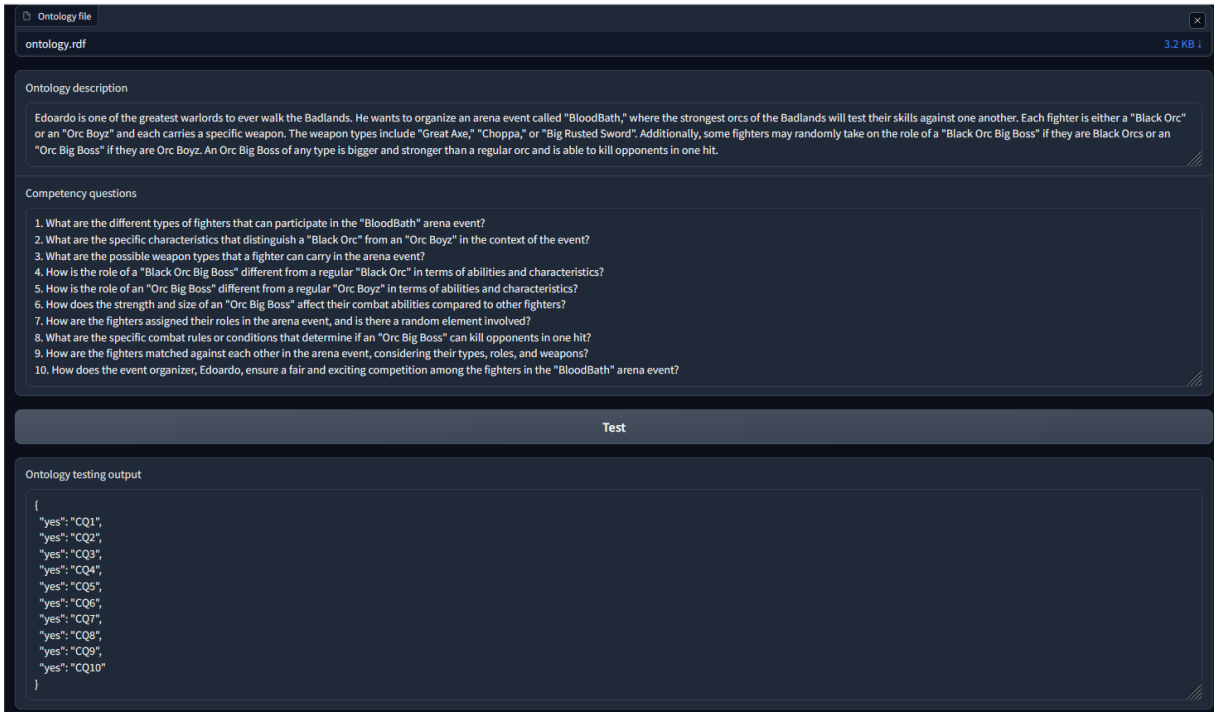

Figure 8

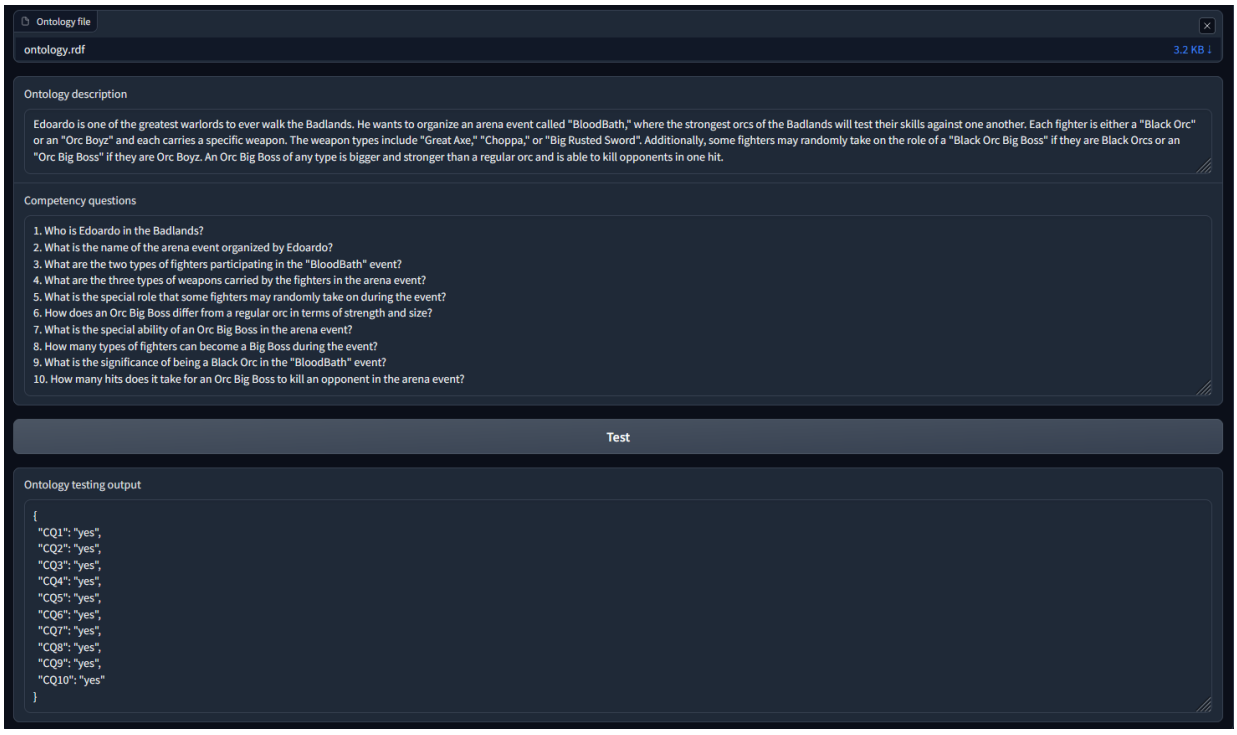

Figure 9

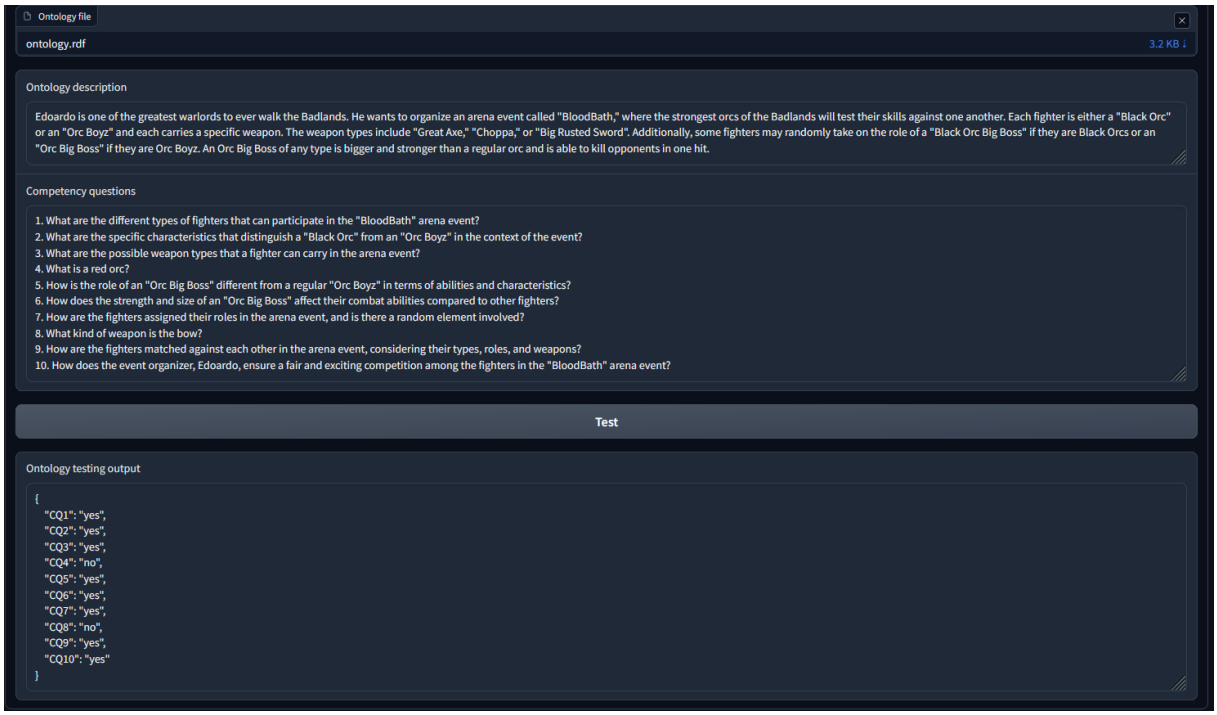

Figure 10

As observed in Figures 8 and 9, the model is able to respond positively to the competency questions (CQs) extracted using both features, as they are closely related to the story. This suggests that the drafted ontology RDF file is robust and can handle different subsets of questions effectively. Furthermore, as shown in Figure 10, the model successfully answers the original questions while responding negatively to the newly introduced, unrelated questions, such as question 4 and question 8:

- *What is a red orc?*
- *What kind of weapon is the bow?*

These results highlight the model's capacity to differentiate between relevant and irrelevant queries, ensuring the integrity of the ontology structure.

4.2 Second Test

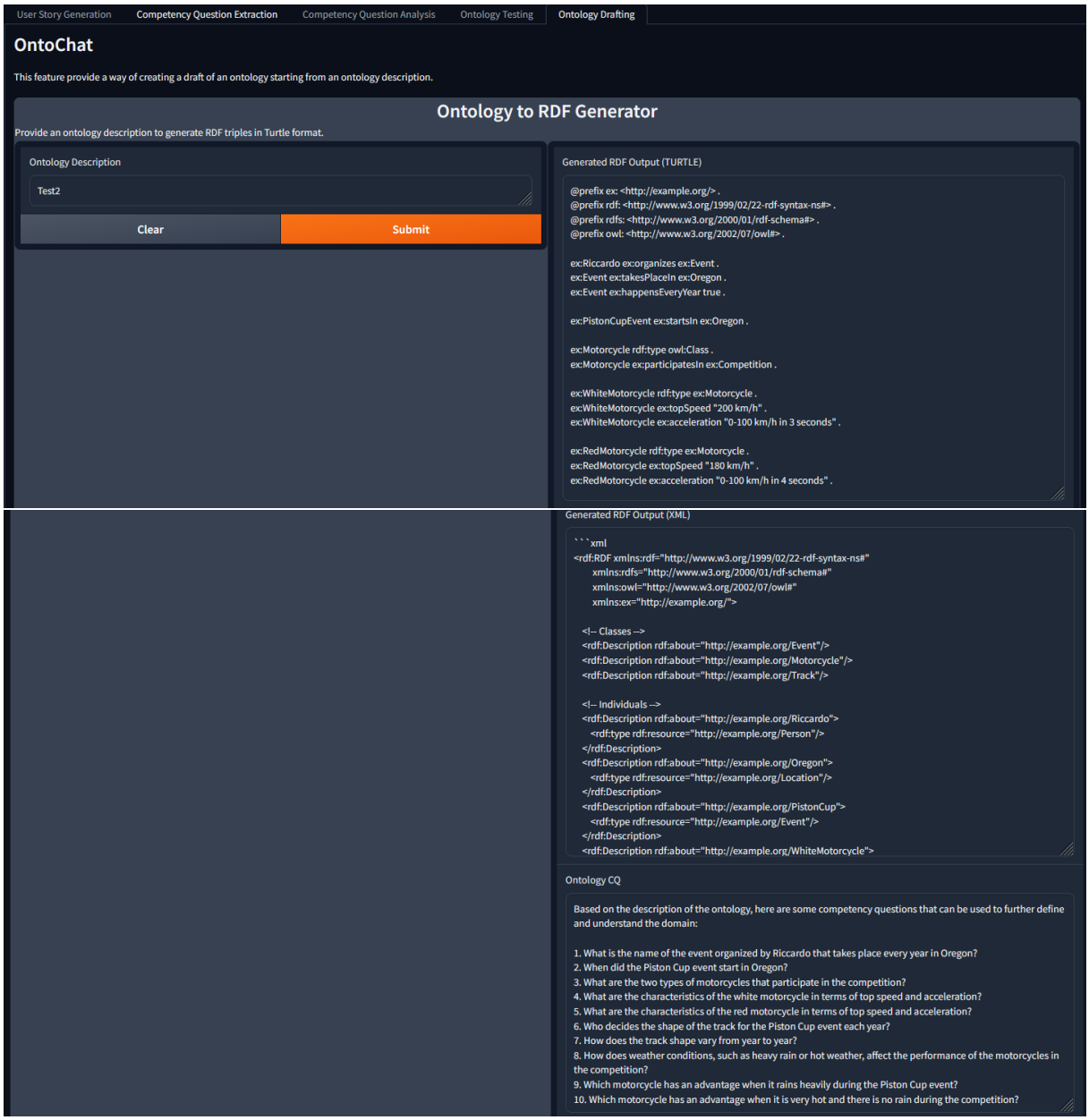

Figure 11

These are the competency questions extracted from the **Ontology_drafting** feature, as seen in figure 11:

1. *What is the name of the event organized by Riccardo that takes place every year in Oregon?*
2. *When did the Piston Cup event start in Oregon?*
3. *What are the two types of motorcycles that participate in the competition?*
4. *What are the characteristics of the white motorcycle in terms of top speed and acceleration?*

5. *What are the characteristics of the red motorcycle in terms of top speed and acceleration?*
6. *Who decides the shape of the track for the Piston Cup event each year?*
7. *How does the track shape vary from year to year?*
8. *How does weather conditions, such as heavy rain or hot weather, affect the performance of the motorcycles in the competition?*
9. *Which motorcycle has an advantage when it rains heavily during the Piston Cup event?*
10. *Which motorcycle has an advantage when it is very hot and there is no rain during the competition?*

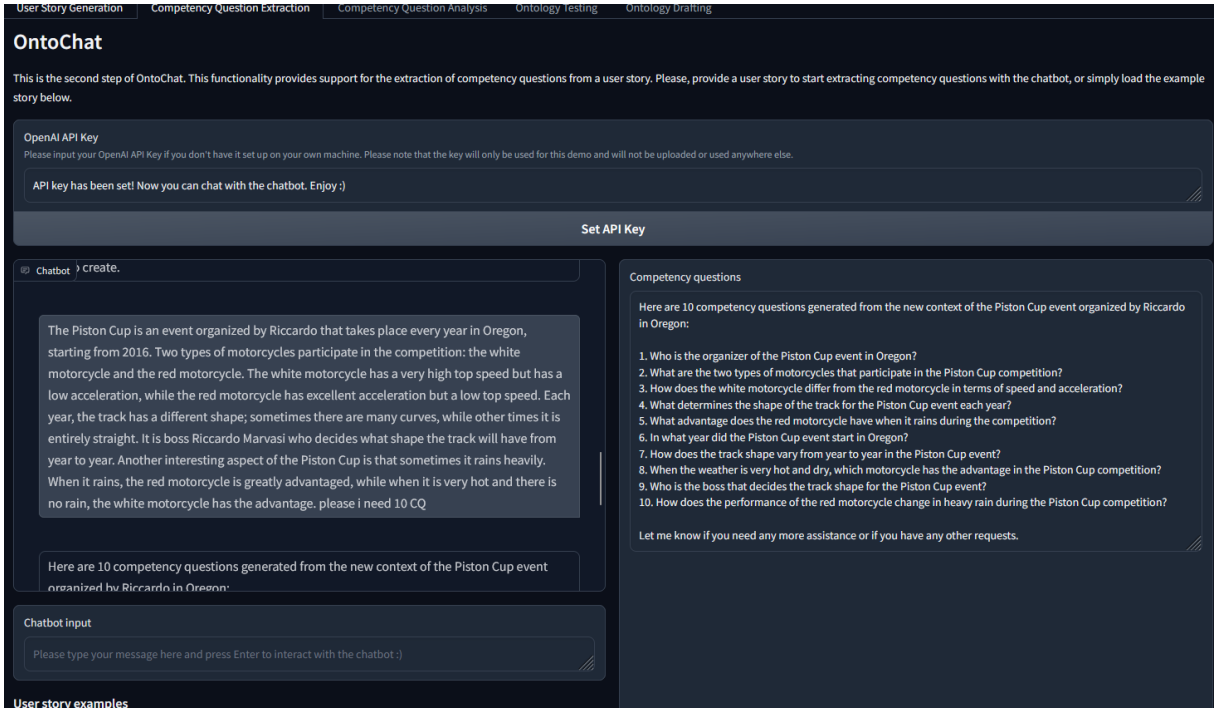

Figure 12

These are the competency question extracted from **Competency_question_extraction** features, as seen in figure 12:

1. *Who is the organizer of the Piston Cup event in Oregon?*
2. *What are the two types of motorcycles that participate in the Piston Cup competition?*
3. *How does the white motorcycle differ from the red motorcycle in terms of speed and acceleration?*
4. *What determines the shape of the track for the Piston Cup event each year?*
5. *What advantage does the red motorcycle have when it rains during the competition?*
6. *In what year did the Piston Cup event start in Oregon?*
7. *How does the track shape vary from year to year in the Piston Cup event?*
8. *When the weather is very hot and dry, which motorcycle has the advantage in the Piston Cup competition?*
9. *Who is the boss that decides the track shape for the Piston Cup event?*

10. How does the performance of the red motorcycle change in heavy rain during the Piston Cup competition?

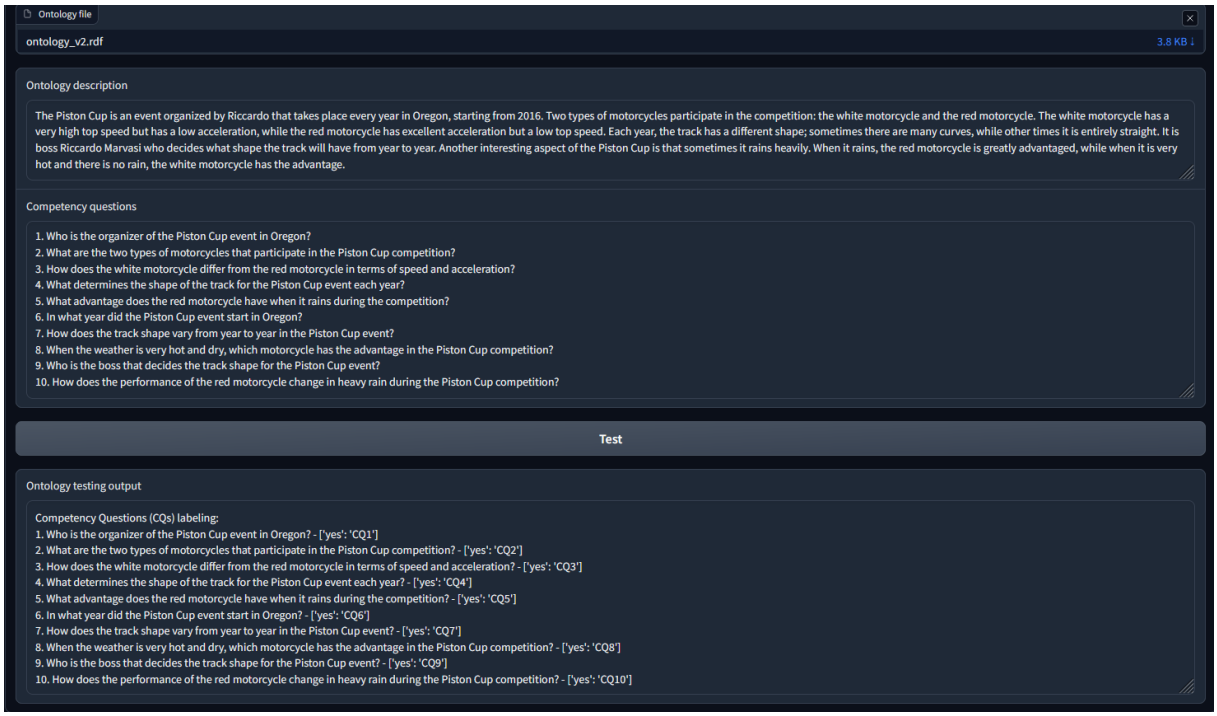

Figure 13

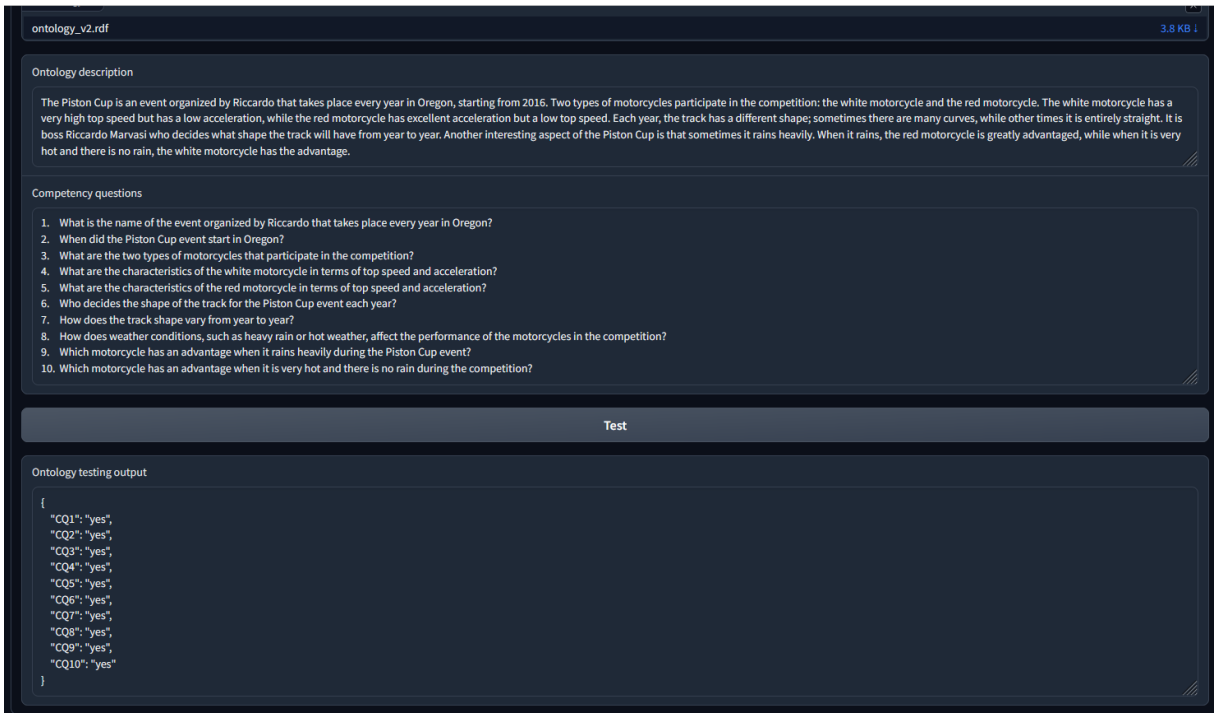

Figure 14

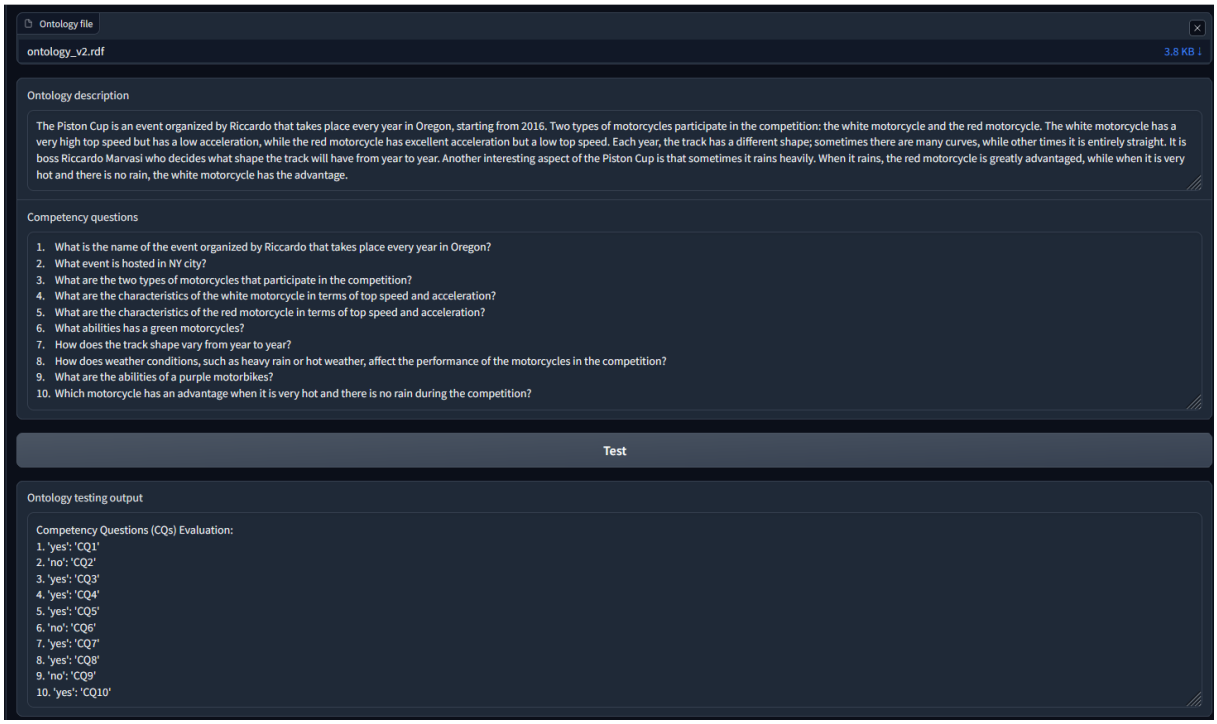

Figure 15

As observed in Figures 13 and 14, the model is able to respond positively to the competency questions (CQs) extracted using both features, as they are closely related to the story. This suggests that the drafted ontology RDF file is robust and can handle different subsets of questions effectively. Furthermore, as shown in Figure 15, the model successfully answers the original questions while responding negatively to the newly introduced, unrelated questions, such as question 2,6 and 9:

- *What event is hosted in NY city?*
- *What abilities has a green motorcycle?*
- *What are the abilities of a purple motorbike?*

These results highlight the model's capacity to differentiate between relevant and irrelevant queries, ensuring the integrity of the ontology structure.

4.3 Third Test Results

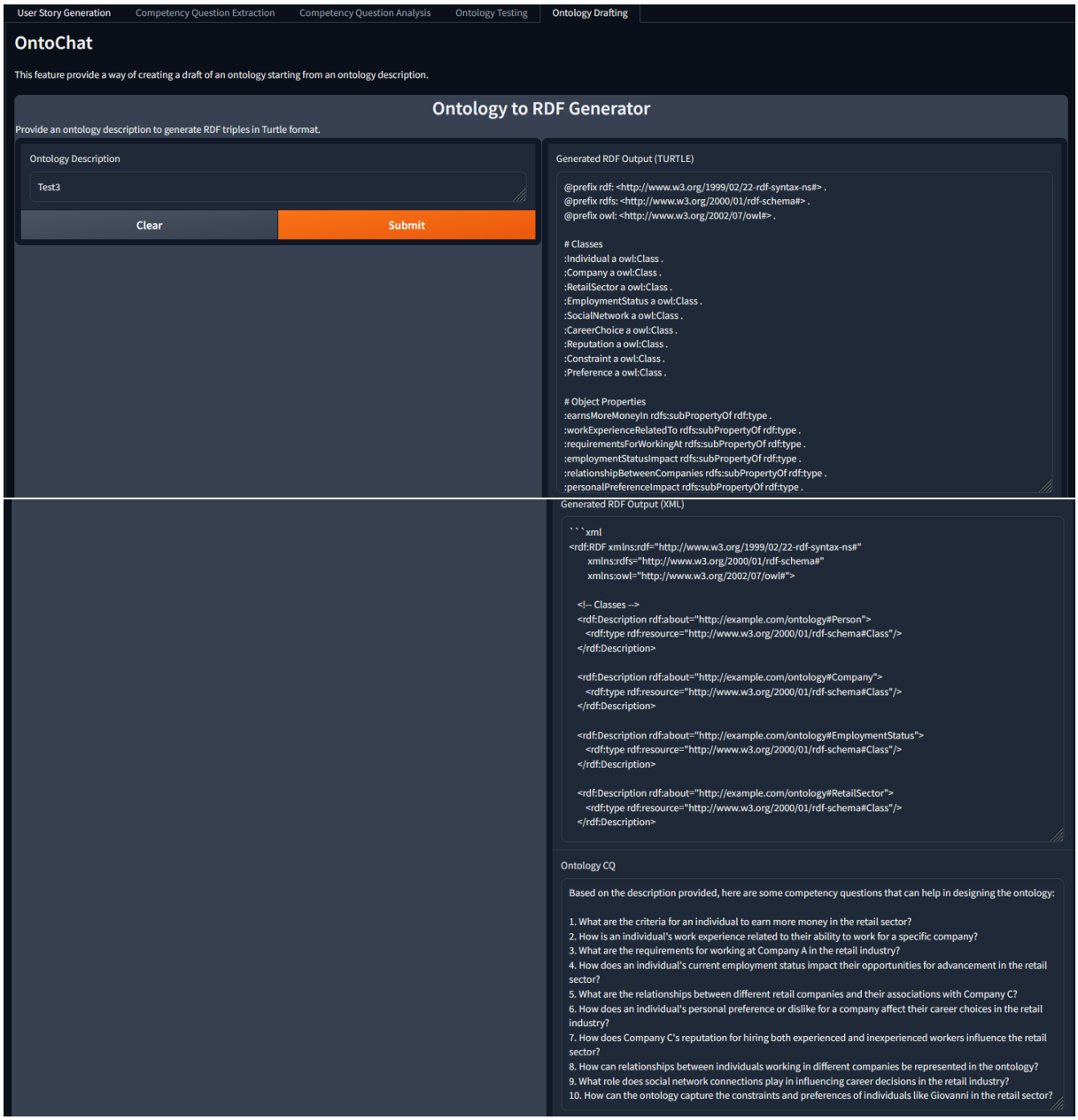

Figure 16

These are the competency questions extracted from **Ontology_drafting** feature, as seen in figure 16:

1. *What are the criteria for an individual to earn more money in the retail sector?*
2. *How is an individual's work experience related to their ability to work for a specific company?*
3. *What are the requirements for working at Company A in the retail industry?*
4. *How does an individual's current employment status impact their opportunities for advancement in the retail sector?*
5. *What are the relationships between different retail companies and their associations with Company C?*

6. *How does an individual's personal preference or dislike for a company affect their career choices in the retail industry?*
7. *How does Company C's reputation for hiring both experienced and inexperienced workers influence the retail sector?*
8. *How can relationships between individuals working in different companies be represented in the ontology?*
9. *What role does social network connections play in influencing career decisions in the retail industry?*
10. *How can the ontology capture the constraints and preferences of individuals like Giovanni in the retail sector?*

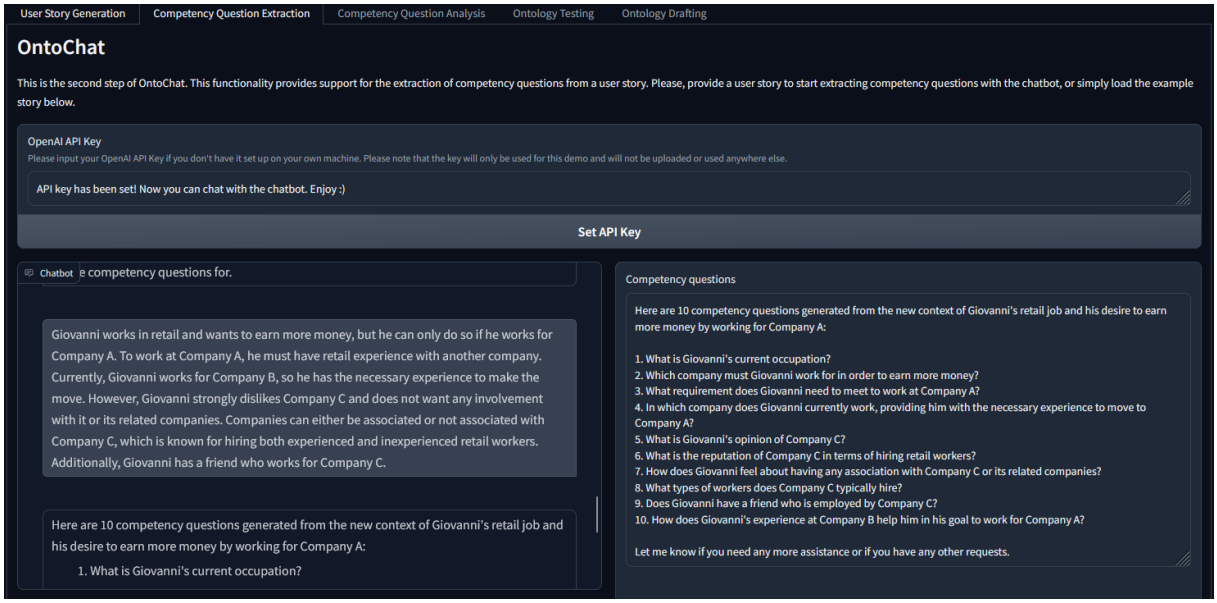

Figure 17

These are the competency questions extracted from **Competency_question_extraction** feature, as seen in figure 17:

1. *What is Giovanni's current occupation?*
2. *Which company must Giovanni work for in order to earn more money?*
3. *What requirement does Giovanni need to meet to work at Company A?*
4. *In which company does Giovanni currently work, providing him with the necessary experience to move to Company A?*
5. *What is Giovanni's opinion of Company C?*
6. *What is the reputation of Company C in terms of hiring retail workers?*
7. *How does Giovanni feel about having any association with Company C or its related companies?*
8. *What types of workers does Company C typically hire?*
9. *Does Giovanni have a friend who is employed by Company C?*
10. *How does Giovanni's experience at Company B help him in his goal to work for Company A?*

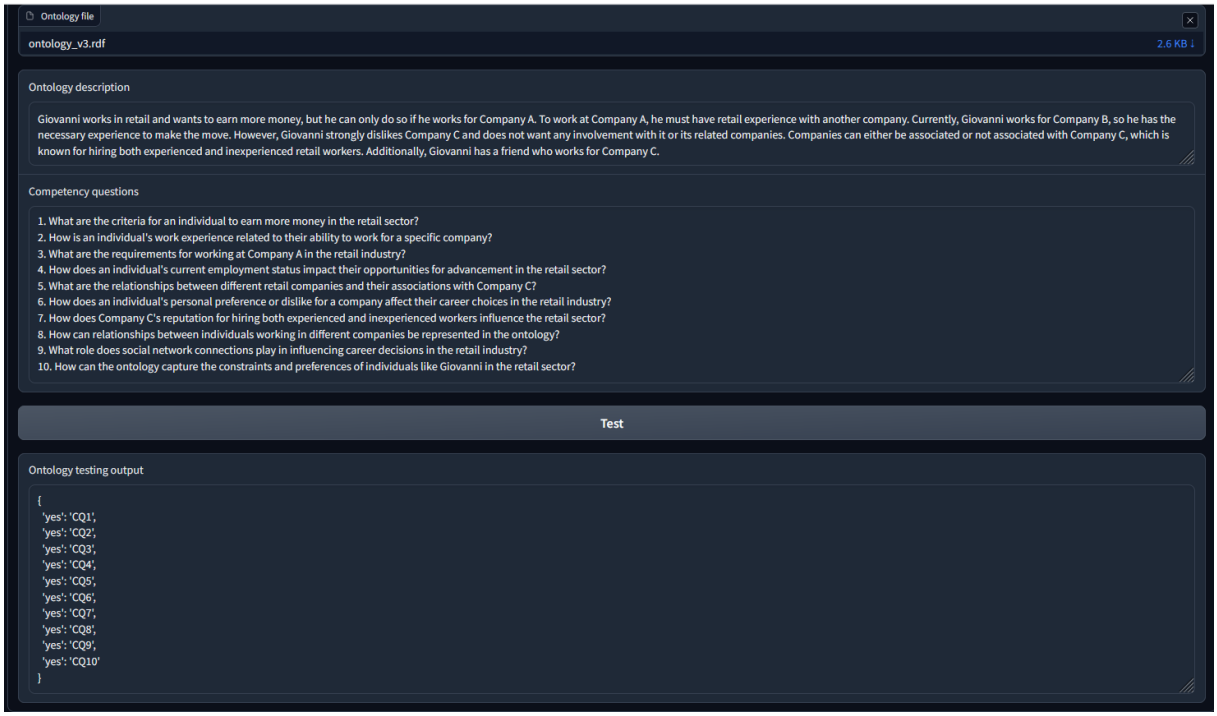

Figure 18

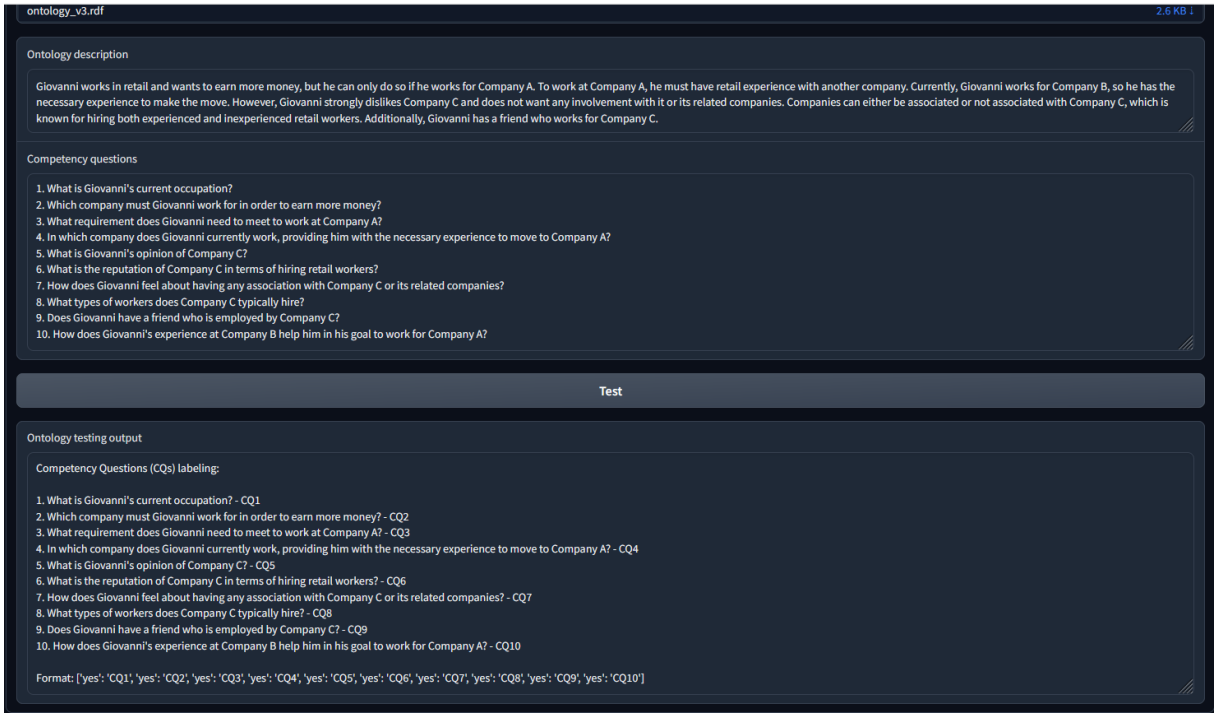

Figure 19

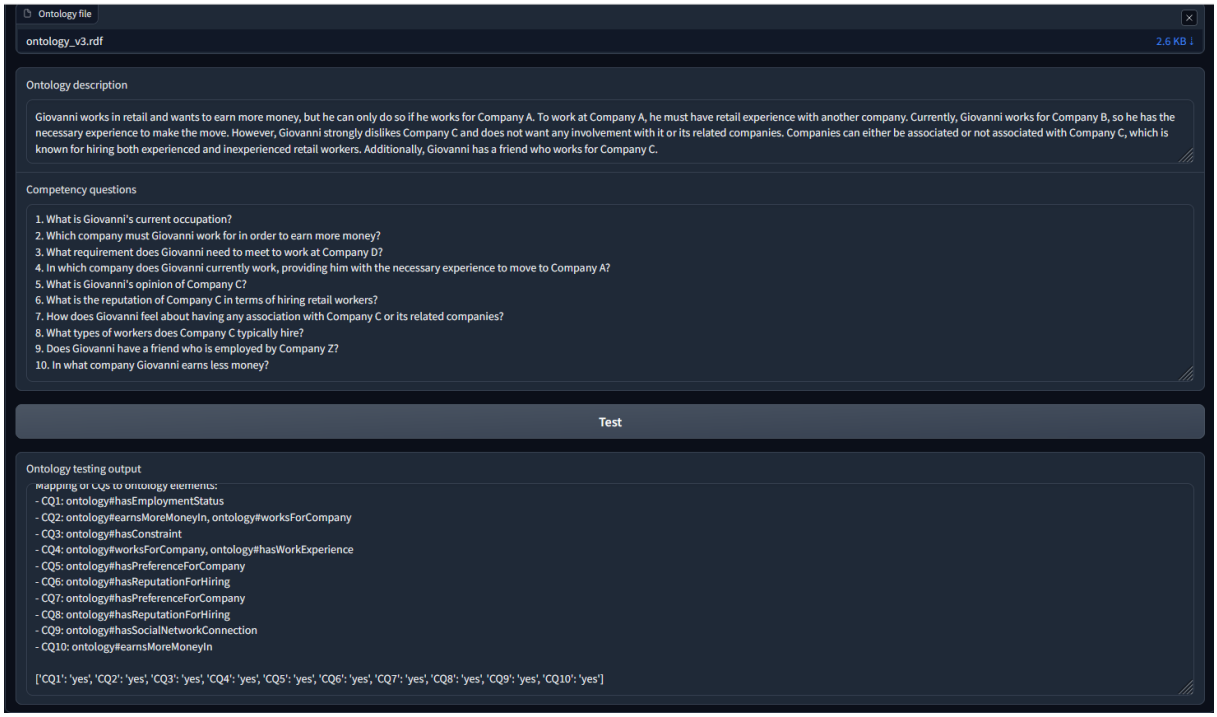

Figure 20

As observed in Figures 18 and 19, the model is able to respond positively to the competency questions (CQs) extracted using both features, as they are closely related to the story. This suggests that the drafted ontology RDF file is robust and can handle different subsets of questions effectively. This time, as shown in Figure 20, the model is not able to respond correctly to the “noise” questions 3,9 and 10:

- *What requirement does Giovanni need to meet to work a company D?*
- *Does Giovanni have a friend that works in company Z?*
- *In what company Giovanni earns less money?*

5 Discussion

In the first two examples, we observed that the model was able to correctly differentiate between related and unrelated questions and generate an RDF file robust enough to handle variations in the questions. This allowed the model to respond appropriately to the competency questions. However, in the third and final example, the model responded positively to all three groups of questions, including the new, unrelated ones. This highlights a misunderstanding of our request by the large language model (LLM). But why does this happen?

It is widely known that LLMs often struggle with basic logic and mathematical reasoning, as demonstrated in a series of experiments conducted by data engineer Thomas Reid [2]. The problem in the third example is that it requires both abstraction and logical reasoning. This time, we generalized the situation by assigning a placeholder name, such as "Z", to a new random company. The model might have mistaken this new company for one it had encountered earlier, leading to incorrect positive responses.

This demonstrates that even with customized improvements to ontology tools, human oversight is still necessary to combine machine output with human understanding, especially in tasks requiring abstraction. Since we are only observing machine-generated results in this case, the model's misunderstanding becomes apparent. This explains the discrepancies seen in the third test case.

6 Conclusion

OntoChat demonstrates significant potential in improving ontology requirements engineering by automating the process of generating user stories, competency questions (CQs), and draft ontologies. Its use of large language models (LLMs) allows for streamlined workflows, making it easier to create and refine ontologies based on real-world user inputs. The system excels in producing robust RDF and Turtle files capable of handling question variations and provides valuable support for ontology testing.

However, the tool has limitations. Our experiments reveal that OntoChat struggles with tasks requiring higher levels of abstraction and logical reasoning. As seen in the discussion, while the model can effectively differentiate between related and unrelated questions in simpler scenarios, it falters when faced with more abstract, generalized cases, such as dealing with placeholder entities. This highlights the known challenges LLMs face with logic and reasoning, as well as the ongoing need for human intervention to ensure the quality and correctness of ontology outputs.

In conclusion, while OntoChat offers a promising framework for ontology engineering, particularly in terms of automating repetitive tasks and facilitating user interaction, its limitations in abstract reasoning and logic suggest that it works best as a collaborative tool, where human expertise complements machine-generated insights.

7 References

1. https://github.com/King-s-Knowledge-Graph-Lab/OntoChat
2. https://medium.com/@thomas_reid/why-cant-llms-do-simple-maths-b509090e774a