Expert Interview – Think-Aloud Protocol

Type 1: Expert follows think-aloud protocol.

Research Project: Caption Generation for Artworks

Introduction

This research focuses on the artwork image captioning task – a process designed to describe and facilitate others' understanding and interpretation of art. We propose a new semantic approach to automate the creation of captions for describing artworks, aiming to reduce the arduous, manual, and time-consuming task that demands expert knowledge from multiple domains.

To help us build a semantic approach and automate the generation of captions, we will present interviewees with examples of artworks and ask them to write captions for up to four paintings using the think-aloud protocol – a process that requires the interviewee to articulate their thought processes from the beginning to the end of the creation process. A series of questions may be asked during the interview to clarify their thought process.

The interview process is described in what follows:

Practice

1. The interviewee will be asked:
   - Write a caption for the following artworks and verbalize your thoughts and reasoning process out loud. (Note: Please write this caption with as much detail as possible, covering every aspect you can think of.)

Portrait of Yasuo Kuniyoshi in His Studio	The Last Supper
The Night Watch	Rysdyk's Hambletonian

The Gleaners	Barge Haulers on the Volga
Paris Street; Rainy Day	The Song of the Lark
The Banquet	The Happy Mother

- After writing the captions for the artworks, the interviewee will be presented with the definitions of different types of captions ('Descriptive', 'Contextual' or 'Interpretative'), as seen below, and will be asked to label each part of the caption following those types.

- Types of Captions:
    - Descriptive: The descriptive caption focuses on detailing what is visually present in the artwork.
    - Contextual: The contextual caption provides background information about the artwork, aiming to place it within its historical and/or cultural contexts.
    - Interpretative: The interpretative caption delves beyond the surface, offering a deeper analysis on the meaning of the artwork. It typically includes the symbolism, thematic, or artistic intent behind the visual components.

2. To understand the thought process of the interviewee, we may ask the following questions:
   - What aspects are you focus on when writing the caption for paintings? Why?
   - Did you follow an order of relevance when describing this painting?
     - For example: #1 – the landscape; #2 – the jar; #3 – blue colour, etc.
   - Would this order have different levels of relevance?
     - For example: #1 – 50%; #2 – 30%; #3 – 5%, etc.
   - Could you explain the methodology used to write the caption?
   - Other follow up questions in case expert does not mention are:
     - Are any of the "placeholder" relevant to the generation of the caption? Could you explain your answer?
     - Items to replace the placeholder above.
       - Artist.
       - Background information about the Artist.
       - Background information about the Painting.
       - Composition.
       - Content.
       - Emotion (from the perspective of the viewer and the artist).
       - Interpretation.
       - Objects.
       - Style / Technique.
       - Theme of the Painting.
       - Visual Elements.
   - In your opinion, what is the purpose of captions?
   - Have you considered a target audience for the generated captions? Who are they? Why?
   - Do you use any process to validate the generated captions? Which one?
   - Is there an ideal length for the captions? How to determine it?
   - What are the most challenging aspects of caption generation?
   - What's the most common mistake you see when novices try to caption artworks?
   - How to avoid bias when generating captions?
   - How do you handle artworks with controversial or potentially sensitive content when generating captions?
   - Is there anything else you would like to discuss besides these questions? What are they? Please feel free to share any further thoughts or insights.

Type 2: ask specific questions. Show them our design and ask them the problem of our design.

Introduction

This research focuses on the artwork image captioning task – a process designed to describe and facilitate others' understanding and interpretation of art. We propose a new semantic approach to automate the creation of captions for describing artworks, aiming to reduce the arduous, manual, and time-consuming task that demands expert knowledge from multiple domains.

To help us build a semantic approach and automate the generation of captions, we will present interviewees with examples of knowledge graphs – a structured representation of knowledge that uses nodes and edges to organize and connect information. These graphs contain

detailed information about artworks, and based on them, we will ask interviewees to write captions for up to four paintings using the think-aloud protocol – a process that requires the interviewee to articulate their thought processes from the beginning to the end of the creation process. A series of questions may be asked during the interview to clarify their thought process.

The interview process is described in what follows:

Practice

- The interviewee will be asked:
  - Write a caption for the painting based on the knowledge graph of the painting.
  - After writing the caption, the interviewee will be presented with the painting described in the graphs and asked about missing information in the graphs.
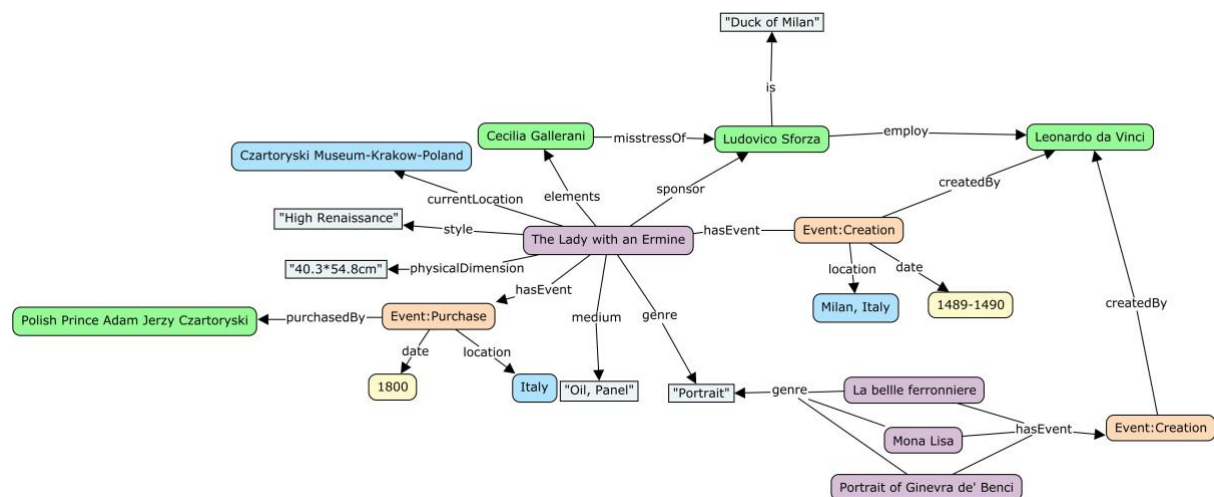
The "context" of the painting "The Lady with an Ermine".
(The "context" includes the basic information (i.e. artist, style, creation date), the historical information and artist's background about the artwork.)
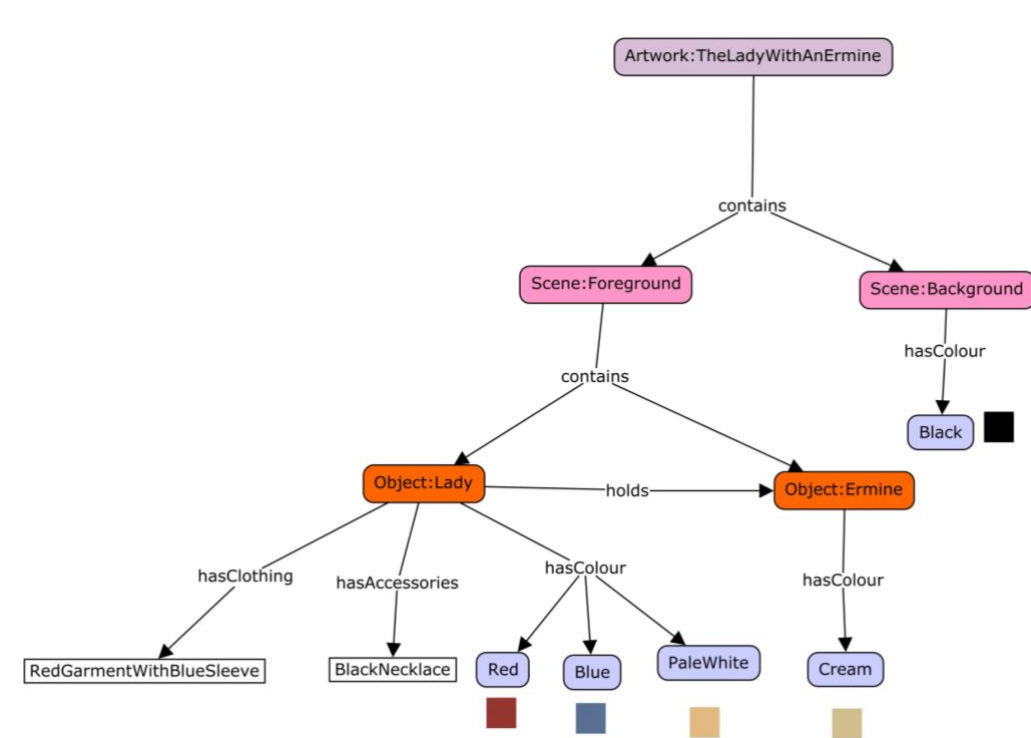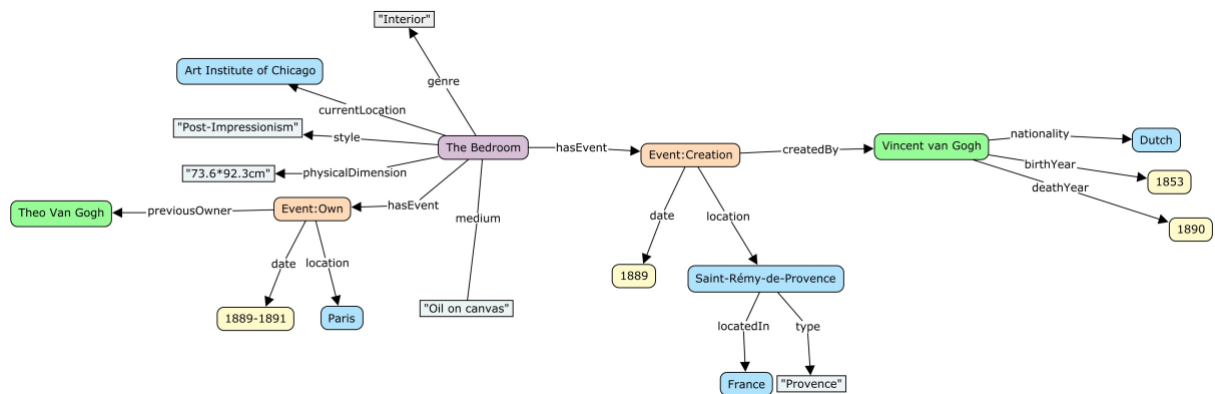
The "content" of the painting "The Lady with an Ermine".
(The "content" refers to the visual component depicted within the painting itself. It includes the basic elements, object, scene and emotion.)
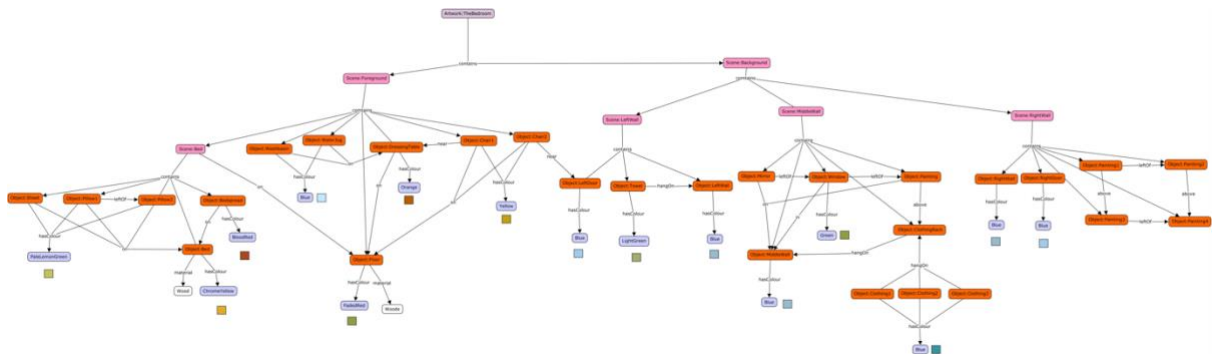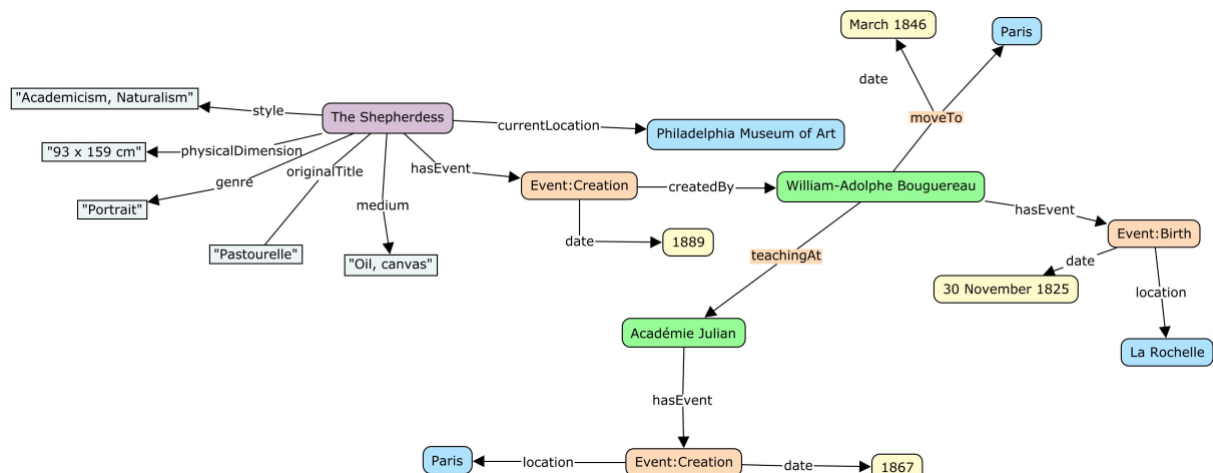

"The Lady with an Ermine"

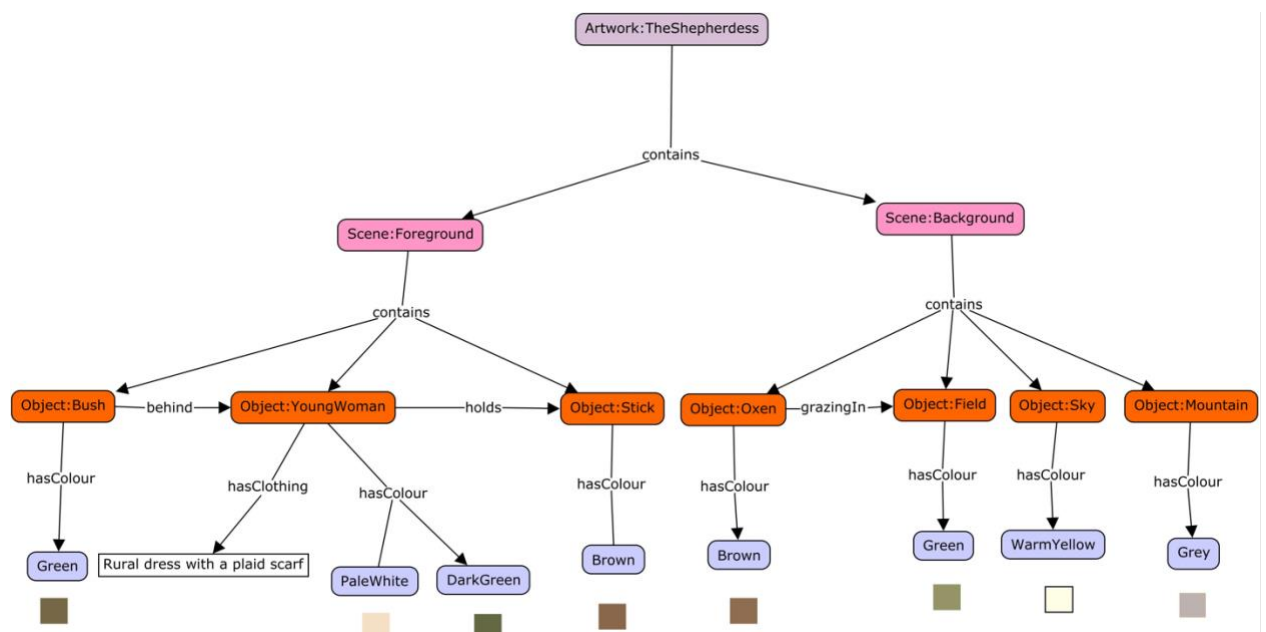The "context" of the painting "The Bedroom".

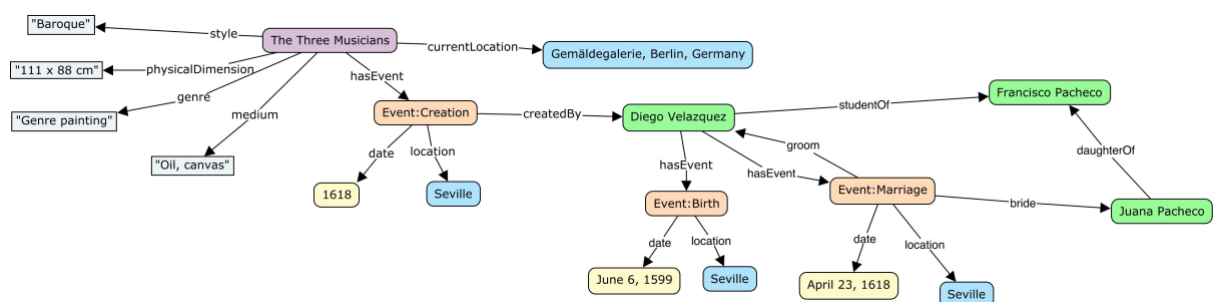The "content" of the painting "The Bedroom".

"The Bedroom"

The "content" of the painting "The Shepherdess".

The "content" of the painting "The Shepherdess".
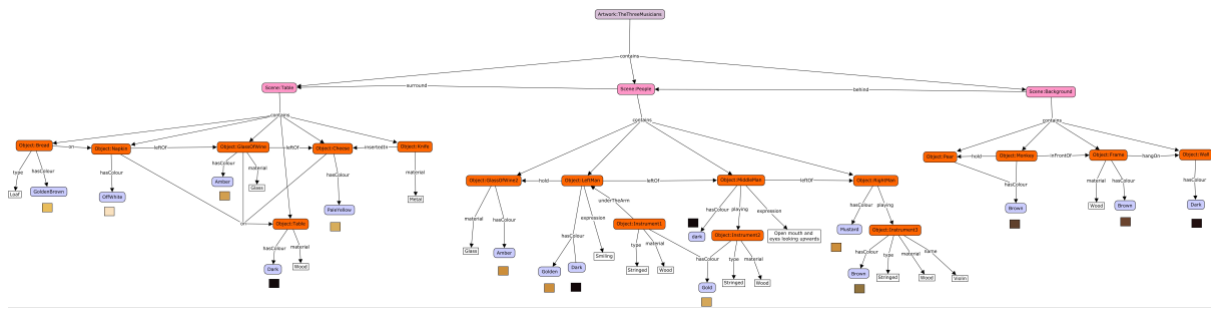
"The Shepherdess"

The "content" of the painting "The Three Musicians".

The "content" of the painting "The Three Musicians".

"The Three Musicians"

General Questions

In this section we will ask you about the details you might consider and the steps or approaches to clarify captioning process.

1. Could you label each part of the caption you generated as "Descriptive" or "Contextual" and "Interpretative"?
   a. Descriptive: The descriptive caption focuses on detailing what is visually present in the artwork.
   b. Contextual: The contextual caption provides background information about the artwork, aiming to place it within its historical and/or cultural contexts.
   c. Interpretative: The interpretative caption delves beyond the surface, offering a deeper analysis on the meaning of the artwork. It typically includes the symbolism, thematic, or artistic intent behind the visual components.
2. Is there any other label you would use to label the captions? Why?

3. Is there any difference between the process of creating captions only based on the graph and based on the painting? What's the difference?
4. In your opinion, what is the purpose of captions?
5. Is there a target audience for the generated captions? Why?
6. Given the subjectivity of themes and emotions in artworks, how do you cope with potential misinterpretations of the artwork?
7. What elements do you consider when expressing the emotion of an artwork?
8. Are background events relevant to the understanding and description of an artwork? How do you use external events to describe an artwork?
9. What aspects or characteristics of an artwork are the most challenging to capture from your perspective?
10. What's the most common mistake you see when novices try to caption artworks?
11. How do you validate the generated captions?
12. How to determine the length of the caption?
13. How to avoid bias when generating captions?
14. How do you handle artworks with controversial or potentially sensitive content when generating captions?
15. Is there anything else you would like to discuss besides these questions? What are they? Please feel free to share any further thoughts or insights.