DIALOGUE-BASED INFORMATION RETRIEVAL FROM IMAGES

P. Hamřík, I, Kopeček, **R. Ošlejšek**, J. Plhák

LAB OF SOFTWARE
ARCHITECTURES AND
INFORMATION SYSTEMS

FACULTY OF INFORMATICS
MASARYK UNIVERSITY

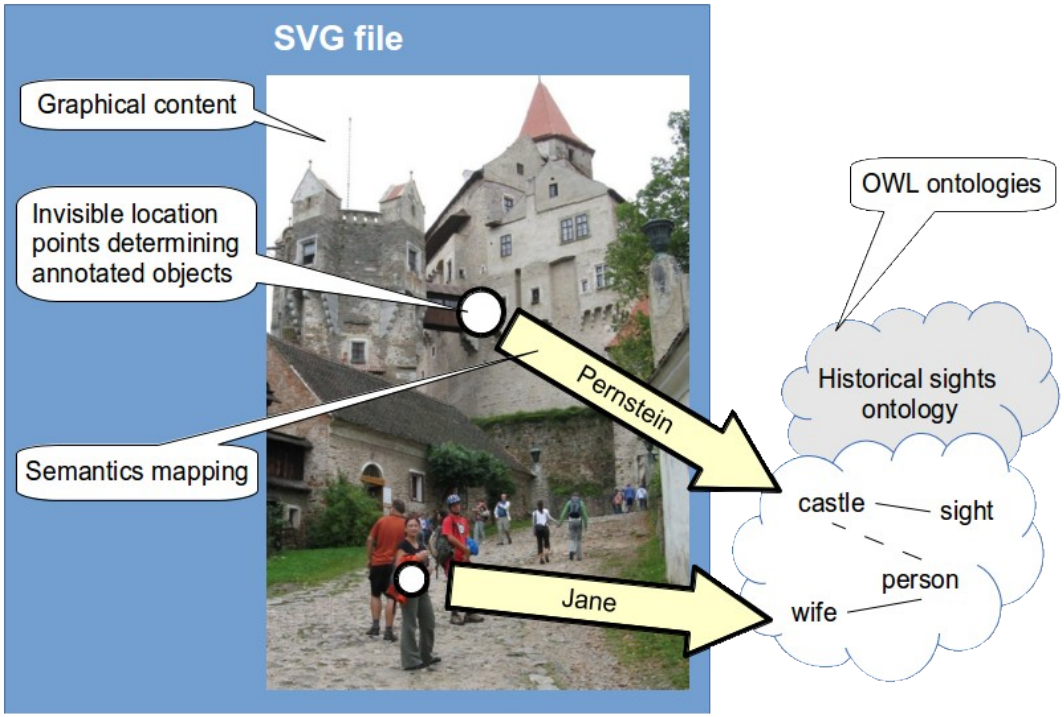
Motivation – Communicative Images

- **Communicative image**
  - An image enabling users to explore its content by means of dialogues.
  - Window to the depicted world fully accessible through natural language.

lasaris

Key Principles – Annotated Pictures

- **Semantics:** System of OWL/RDF ontologies for picture annotation and shared multilingual knowledge. Defines grammar of the dialogue system.
- **Graphic format:** SVG as flexible XML wrapper enabling us to embed the original raster image together with structured semantics



lasaris

Key Principles – Dialogue Subsystem

- **Restricted grammars** (only a small fragment of natural language):
  - Generic grammar: *“Describe picture.”*, *“What is in the picture”*, etc.
  - What-Where Language: *“Where is object?”*, *“What is in the upper-left corner?”*.
  - Experimental domain-specific grammars: Fine-tuned for concrete picture.
- **Dialogue frames:** templates for questions with slots that can be filled by specific entries from ontologies.
  - *“How far is it from SLOT1 to SLOT2?”*

lasaris

Workflow

- Client (e.g. plug-in to web browser) sends an image to our server
  - The image can be communicative or not
  - JavaEE server providing REST services
- Server embeds the image in SVG and do additional preprocessing
  - Auto-detection and image-recognition techniques would help to gather initial semantics [in development]
- Client sends questions (sentences) to the server, dialogue module parses the questions, inspects ontology and composes answer
  - No intelligence on the client side

lasaris

Demo

- Exact transcription of our experimental system
- Adjusted for screen readers:
  - **http://lsd.fi.muni.cz/~xplhak/gate/**
- With styles:
  - **http://lsd.fi.muni.cz/~xplhak/virtualgate/**

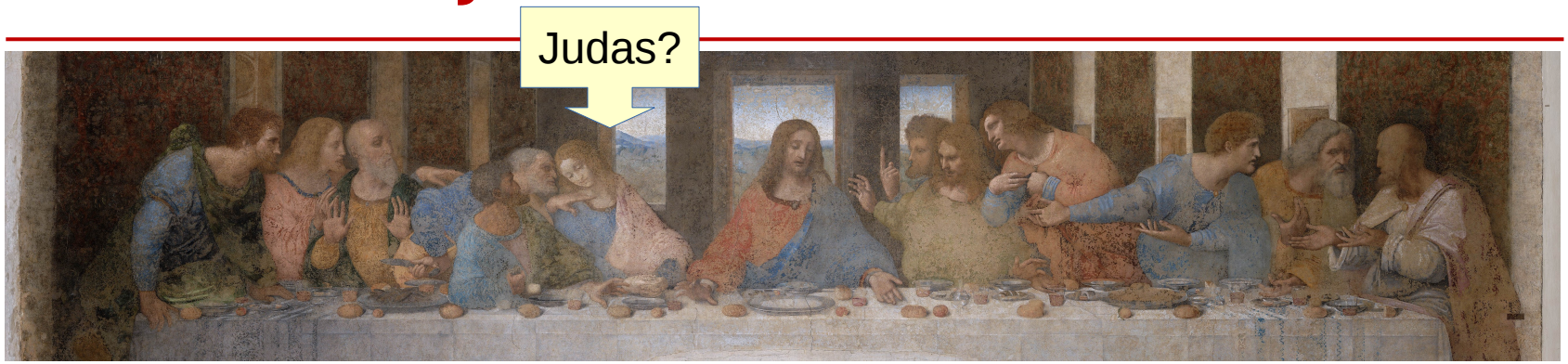
lasaris

Goal 1: identify (hidden) image

- What is in the middle?
- There are Philip, Jesus, Table, John, Peter, Thomas, Judas and James Major in the central part.
- What is Jesus doing?
- In this picture, Jesus is Sitting.
- Hwo is author?
- This painting was created by Leonardo da Vinci.
- When did Leonardo create the painting?
- Painting has been finished in 1498.

lasaris

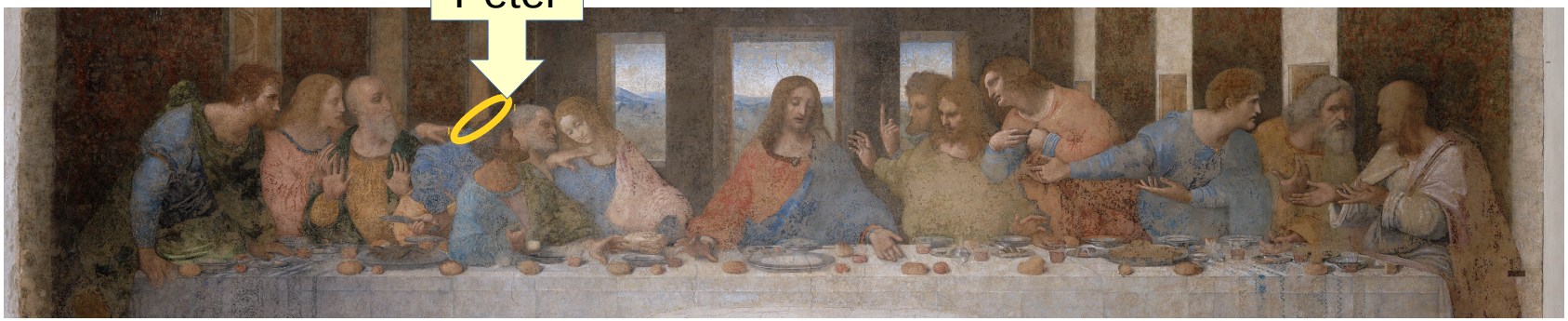
Goal 2: Find Judas



- Where is Judas?
- Judas is in the middle of the picture.
- How Judas looks?
- In this picture, Judas is Withdrawn.
- Is Judas next to Jesus?
- From your point of view, Judas is not next to Jesus. Judas is on the left of Jesus.

lasaris

Goal 2: Find Judas (cont.)



- What is Peter doing?
- In this picture, Peter is Talking and Standing.
- Is Judas in front of Peter?
- From your point of view, Judas is in front of Peter.

lasaris

User evaluation

- 4 blind users and 4 sighted users
- Testing scenarios
  - Start the interaction with the picture in any way you like. And end it at any point you like.
  - If the user haven’t done it in the previous scenario, then:
    - Obtain general information about the picture
    - Learn who painted the painting in the picture.
    - List all people in the picture.
    - ...
- Evaluation: quantitative and qualitative questionnaire

lasaris

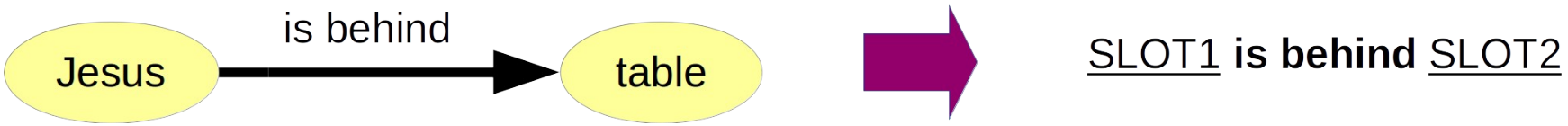
Current Limits and Future Goals

- **Manual annotation**
  - Boring and exhausting, prone to errors even when using supporting tools like Protege.
- **Auto-learning dialogue strategy**
  - User question *“What is the castle behind Jane?”* indicates that there is some castle and some object called Jane in the picture.
  - The communicative picture takes over the initiative to learn more about these two things, asking the user *“Who or what is Jane?”* and then extending the ontology with these new facts.

lasaris

Current Limits and Future Goals (cont.)

- **Manually configured dialogues**
  - Carefully prepared and fine-tuned grammars and dialogue frames for concrete domain (picture content).
- **Dialogues generated from ontologies**
  - Frames driven by ontology structure
  - Object and data properties = frames (utterances).
  - Classes and datatypes involved in properties = slots.
  - Individuals = slot values.



Questions?

Thank you for your attention

lasaris