Multimedia Information Extraction and Retrieval

Multimedia Interpretation

Ralf Moeller
Hamburg Univ. of Technology

Media Interpretation

Possible interpretation:

- "Two workers empty garbage containers into a garbage truck"
- "A mailman distributes mail"

What is a possible high-level annotation for this image?

What do we need?

- Multimedia interpretation: from low-level to high-level (and back!)
- Low-level w.r.t. images: information in pixels
- Mid-levels w.r.t. images: several temporary steps
- High-level w.r.t. images: complex scene description
- Need of representation means for all levels
- $\Rightarrow$ Ontologies can do!

Ontologies

- Tbox
- Abox
  - a:C / C(a) … $a \in C$
  - (a,b):R / R(a,b) … $(a,b) \in R$
- Rules (Datalog)

Interpretations = Metadata (1)

$mailman_1$	:	*Mailman*
$bicycle_1$	:	*Bicycle*
$mail\_deliv_1$	:	*MailDeliv*
$(mail\_deliv_1, mailman_1)$	:	*hasPart*
$(mail\_deliv_1, bicycle_1)$	:	*hasPart*
$(mail\_deliv_1, url\_1)$	:	*hasURL*
$(mailman_1, url\_2)$	:	*hasURL*
$(bicycle_1, url\_3)$	:	*hasURL*
(url_1)	:	$=$`"http://www.img.de/image-1.jpg"`
(url_2)	:	$=$`"http://www.img.de/image-1.jpg#(200,400)/`
(url_3)	:	$=$`"http://www.img.de/image-1.jpg#(100,400)/`

Interpretations = Metadata (2)

$$
\begin{array}{rcl}
 & \ldots & \\
garbageman_1 & : & Garbageman \\
garbageman_2 & : & Garbageman \\
garbagetruck_1 & : & Garbage\_Truck \\
gc_1 & : & Garbage\_Collection \\
(gc_1, garbageman_1) & : & hasPart \\
(gc_1, garbageman_2) & : & hasPart \\
(gc_1, garbagetruck_1) & : & hasPart \\
(gc_1, url\_4) & : & hasURL \\
 & \ldots &
\end{array}
$$

Querying metadata

- Queries:

$$ImageQuery_1 := \{(X,Y) \mid MailDeliv(X), Bicycle(Y), hasPart(X,Y)\}$$
$$URLQuery_1 := \{(X, value(X)) \mid hasURL(mail\_deliv_1, X)\}$$
$$URLQuery_2 := \{(X, value(X)) \mid hasURL(bicycle_1, X)\}$$

- Results :
  - { $(mail\_deliv_1, bicycle_1)$ }
  - { $(url_1$, "http://www.img.de/image-1.jpg") }
  - { $(url_3$, "http://www.img.de/image-1.jpg#(100,400)/(150/500)) }

- Exploit Tboxes: $Mailman \sqsubseteq Postal\_Employee$

$$Mailman \equiv Postman$$

How to derive metadata?

Image analysis

- Assumption: Result of image analysis represented as an Abox

$$pole_1 : Pole$$
$$human_1 : Human$$
$$bar_1 : Bar$$
$$(bar_1, human_1) : near$$

Querying for a pole vault?

- Image will not be found
- Reason:
  - Pole vaults are not directly visible
  - There is no Abox individual for a pole vault
- Idea:
  - Pole vaults are “constructed” to explain the spatial configuration of a person, a bar, and a pole

Explanation via abduction

- Compute some set of assertions Δ such that

$$\Sigma \cup \Delta \models \Gamma$$

- The knowledge base $\Sigma = (\mathcal{T}, \mathcal{A})$
- The set of observables

Γ is a given set of low-level assertions

In our setting

- Image analysis leads to an Abox

$$\begin{aligned} pole_1 &: Pole \\ human_1 &: Human \\ bar_1 &: Bar \\ (bar_1, human_1) &: near \end{aligned}$$

- The last assertion should be explained

$$\Sigma \cup \Gamma_1 \cup \Delta \models \Gamma_2$$

- The Abox is split into parts Γ_1 and Γ_2

Side conditions

- Minimality
- Consilience

Example Tbox

$$
\begin{aligned}
Jumper &\sqsubseteq Human \\
Pole &\sqsubseteq Sports\_Equipment \\
Bar &\sqsubseteq Sports\_Equipment \\
Pole \sqcap Bar &\sqsubseteq \bot \\
Pole \sqcap Jumper &\sqsubseteq \bot \\
Jumper \sqcap Bar &\sqsubseteq \bot \\
Jumping\_Event &\sqsubseteq \exists_{\leq 1} hasParticipant.Jumper \\
Pole\_Vault &\sqsubseteq Jumping\_Event \sqcap \exists hasPart.Pole \sqcap \exists hasPart.Bar \\
High\_Jump &\sqsubseteq Jumping\_Event \sqcap \exists hasPart.Bar \\
near(Y,Z) &\leftarrow Pole\_Vault(X), hasPart(X,Y), Bar(Y), \\
&\quad\; hasPart(X,W), Pole(W), hasParticipant(X,Z), Jumper(Z) \\
near(Y,Z) &\leftarrow High\_Jump(X), hasPart(X,Y), Bar(Y), \\
&\quad\; hasParticipant(X,Z), Jumper(Z)
\end{aligned}
$$

Abduction as query answering

- Abduction equation

$$\Sigma \cup \Gamma_1 \cup \Delta \models \Gamma_2$$

- Entailment of assertions as answering boolean queries w.r.t. Tboxes and Aboxes

$$Q_1 := \{() \mid near(bar_1, human_1)\}$$

Abductive query answering

- The answer should be true!
- What must be added to the Abox to ensure this?
- Idea: Look at the rules
  - Assumption: The rules define the set of abducibles
  - Apply the rules in a backward-chaining way

Backward-chaining

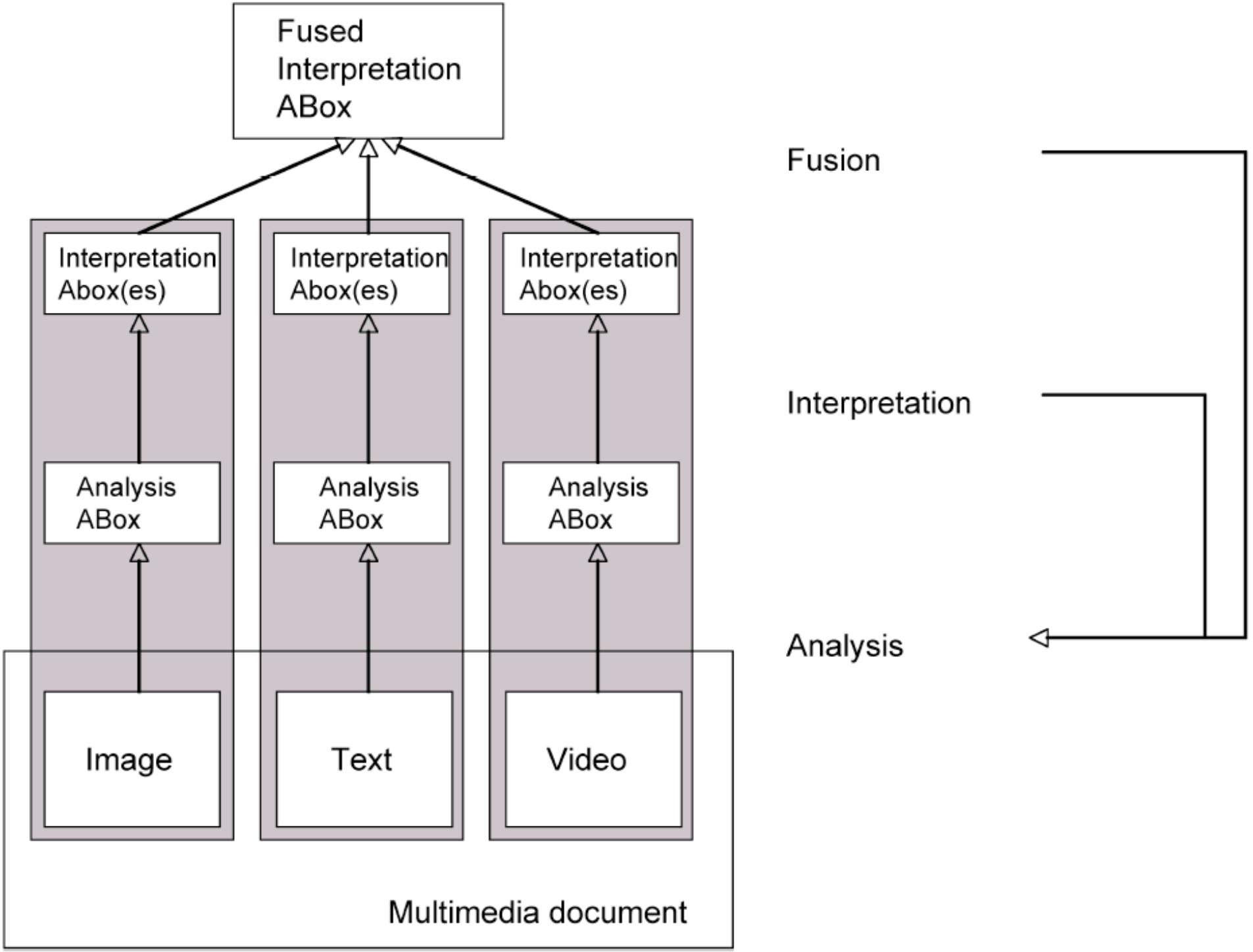
- $$near(Y,Z) \leftarrow Pole\_Vault(X), hasPart(X,Y), Bar(Y), hasPart(X,W), Pole(W), hasParticipant(X,Z), Jumper(Z)$$
  $$near(Y,Z) \leftarrow High\_Jump(X), hasPart(X,Y), Bar(Y), hasParticipant(X,Z), Jumper(Z)$$
- $$Q_1 := \{() \mid near(bar_1, human_1)\}$$
- Determine a substitution for variables
  - Match HEAD with the query
- Expand rules
  - Check all combinations for unbound variables (including new individuals)
  - Role of the Tbox: Discard inconsistent Aboxes obtained by substitutions
- No recursion allowed

Example

- $\Delta_1 = \{new\_ind_1 : Pole\_Vault, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, new\_ind_2) : hasPart,\ new\_ind_2 : Pole,\ (new\_ind_1, human_1) : hasParticipant,\ human_1 : Jumper\}$
- $\Delta_2 = \{new\_ind_1 : Pole\_Vault,\ (new\_ind_1, bar_1) : hasPart,\ (new\_ind_1, pole_1) : hasPart,$
  $(new\_ind_1, human_1) : hasParticipant,\ human_1 : Jumper\}$
- $\Delta_3 = \{new\_ind_1 : High\_Jump, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, human_1) : hasParticipant,$
  $human_1 : Jumper\}$

Preference measures

- When more than one explanation is found, the criteria for selecting the best explanation are:
  - Simplicity
    - $S_s(\Delta) = | \Delta |$
  - Consilience
    - $S_c(\Delta) = | \{ \alpha \in \Gamma_1 : O \cup \Delta \models \alpha \} |$
- In total
  - $S_\omega(\Delta) := \omega S_c(\Delta) - (1 - \omega) S_s(\Delta)$

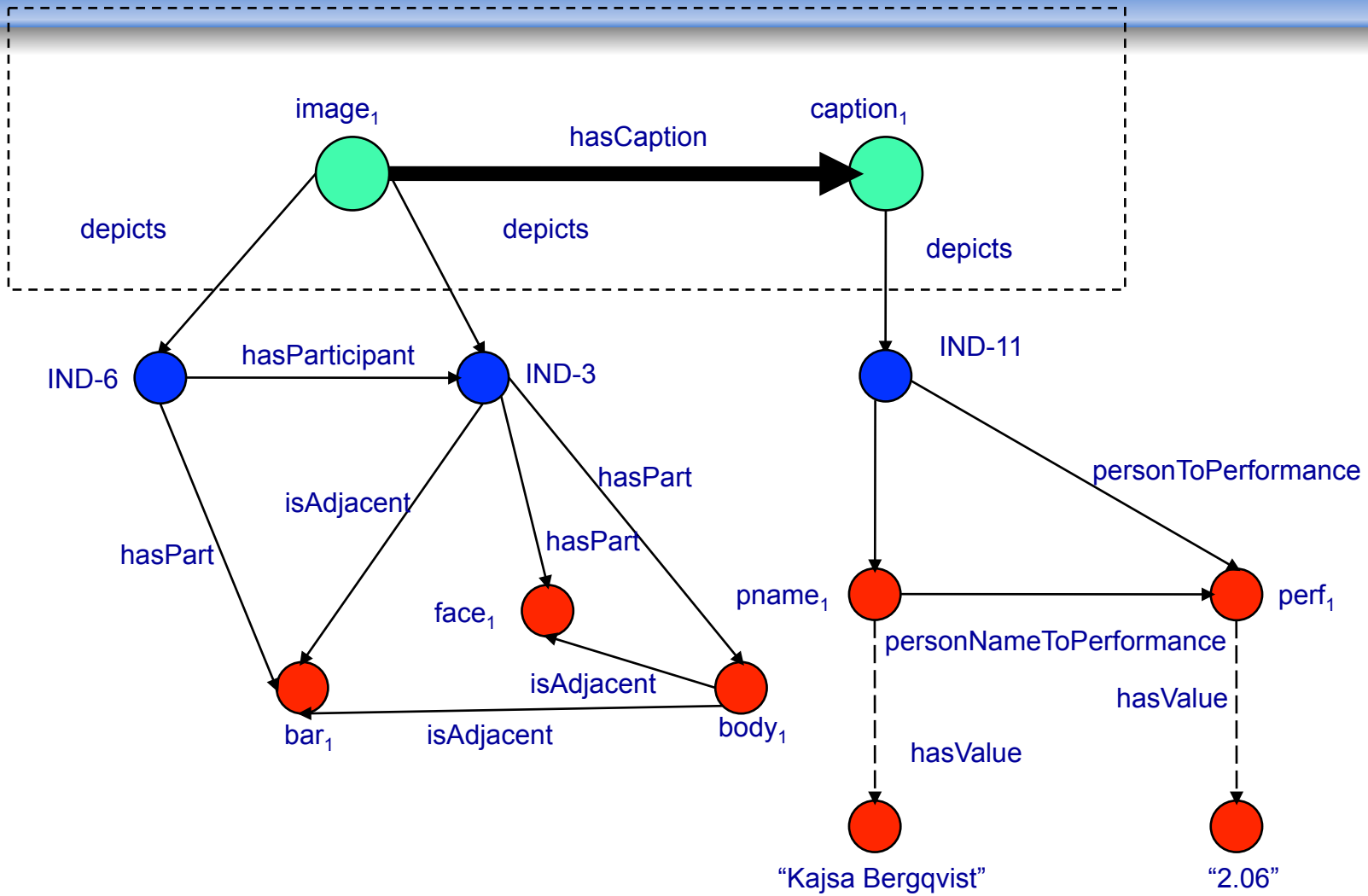

Fused
Interpretation
ABox
Interpretation
Abox(es)
Interpretation
Abox(es)
Interpretation
Abox(es)
Analysis
ABox
Analysis
ABox
Analysis
ABox
Image
Text
Video
Multimedia document
Fusion
Interpretation
Analysis

Fusion

- Goals:
  - Disambiguation
  - Rule out possible interpretations
- Information accumulation (for better query answering)

Kajsa Bergqvist clears 2:06 in Eberstadt

Explicitly represent the document structure

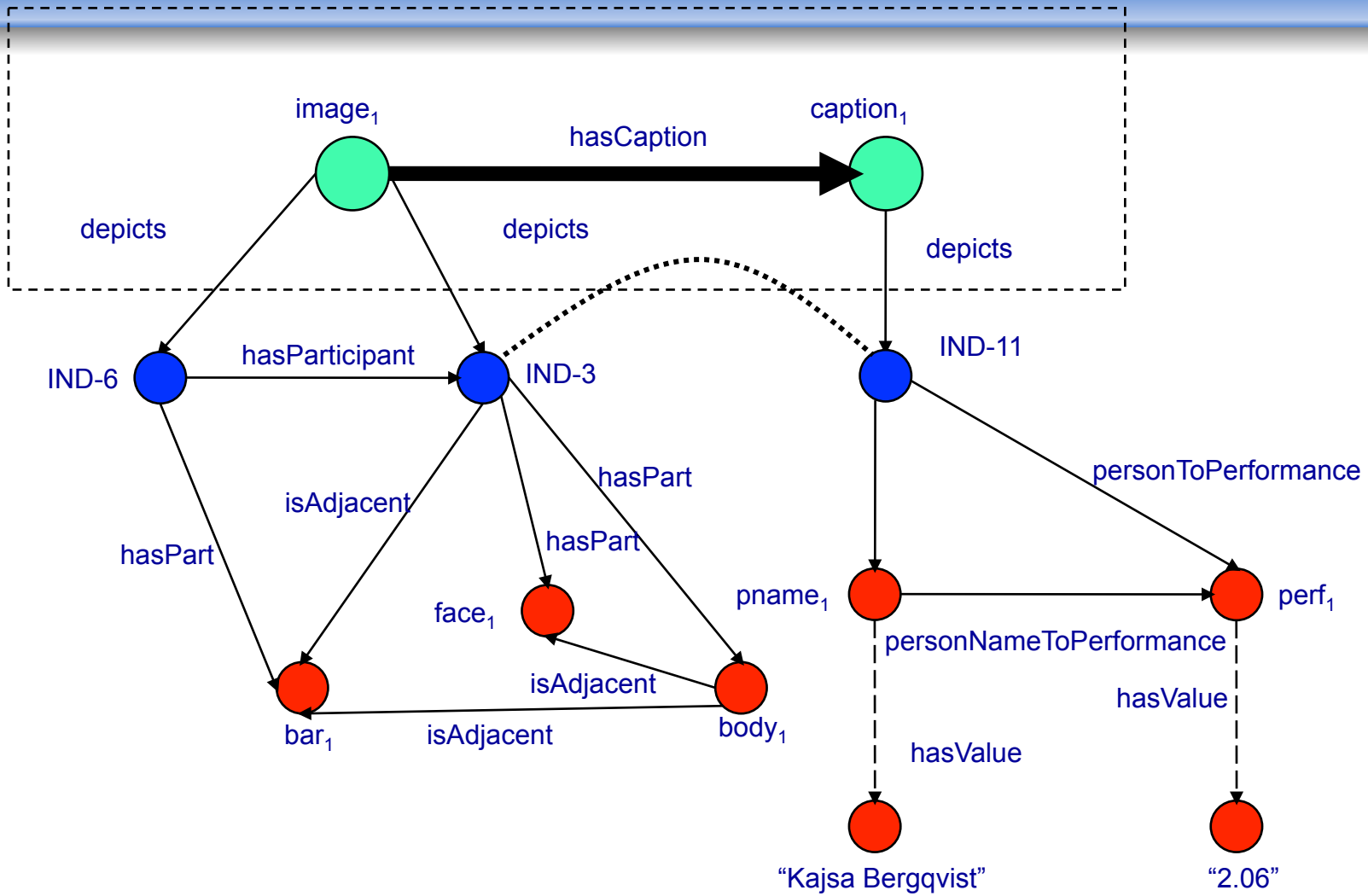


Exploit the document structure

- Abduction is used to find explanations for the relations between the multimedia objects
- We assume the following rule

```
hasCaption(X,A) :-
        Image(X),
        depicts(X,Y),
        Caption(A),
        depicts(A,B),
        same-as(Y,B)
```

Structure identification



Also: Intra-modality fusion

- Example (two sentences):
  - “This car has always been in a garage.”
  - “It always worked well.”
- Disambiguation required:
  - It is the car that worked well, not the garage
  - The car is indeed a good one (not a one that always was in a repair shop)
- Fusion:
  - “This car” = car_1
  - “It” = thing_1
  - Gives rise to abducing car_1 = thing_1

Acknowledgements

- Thanks to the TUHH members of the BOEMIE project (Bootstrapping Ontology Evolution for Multimedia IntEpretation)
  - Sofia Espinosa Peraldi
  - Atila Kaya
  - Sylvia Melzer
- Thanks to Bernd Neumann, UniHH
- Thanks to Michael Wessel