

Multimedia Information Extraction and Retrieval

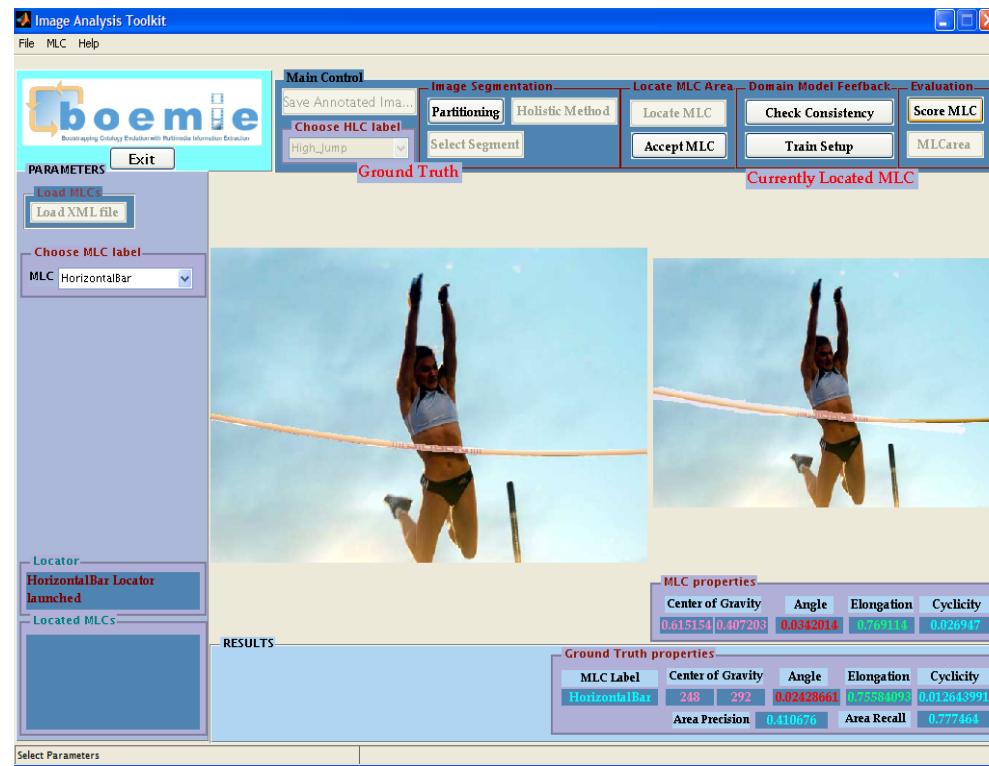
Multimedia Analysis

Ralf Moeller
Hamburg Univ. of Technology

Information extraction from images

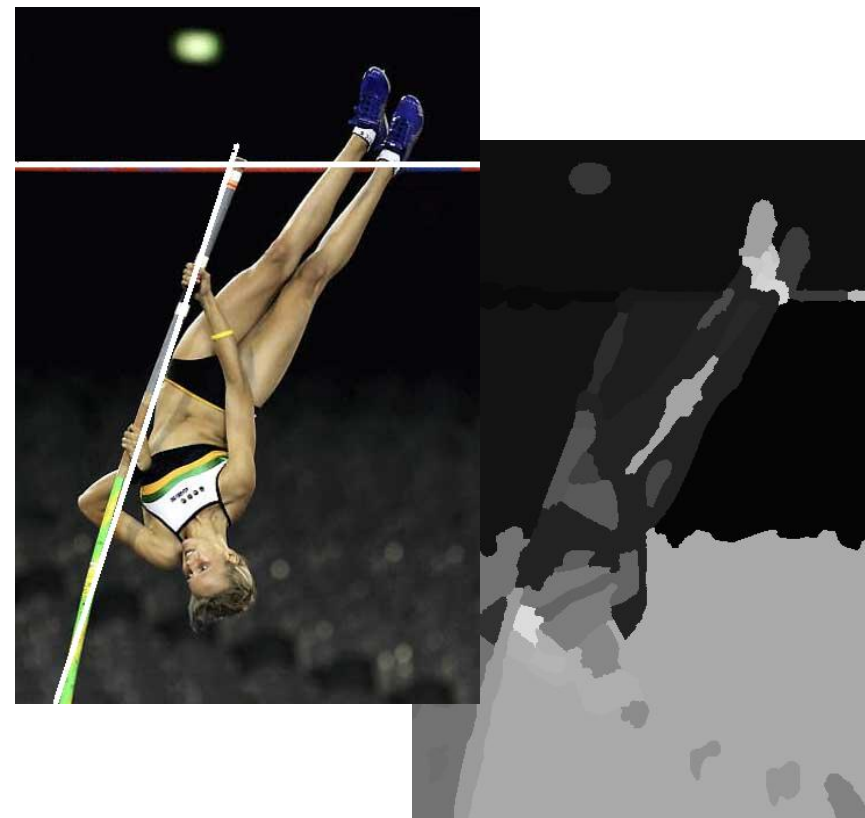
Aim:

- Identify mid-level concept instances (**MLCIs**) (an attach semantic labels: 'pole', 'personbody', 'personface', 'horizontalbar')
- Produces region maps (unique region numbers that identify the **image area** that is covered by a particular mid-level concept instance)
- Determine spatial relations between the regions ('up', 'down' etc)
- Complementary information about **unknown** image regions



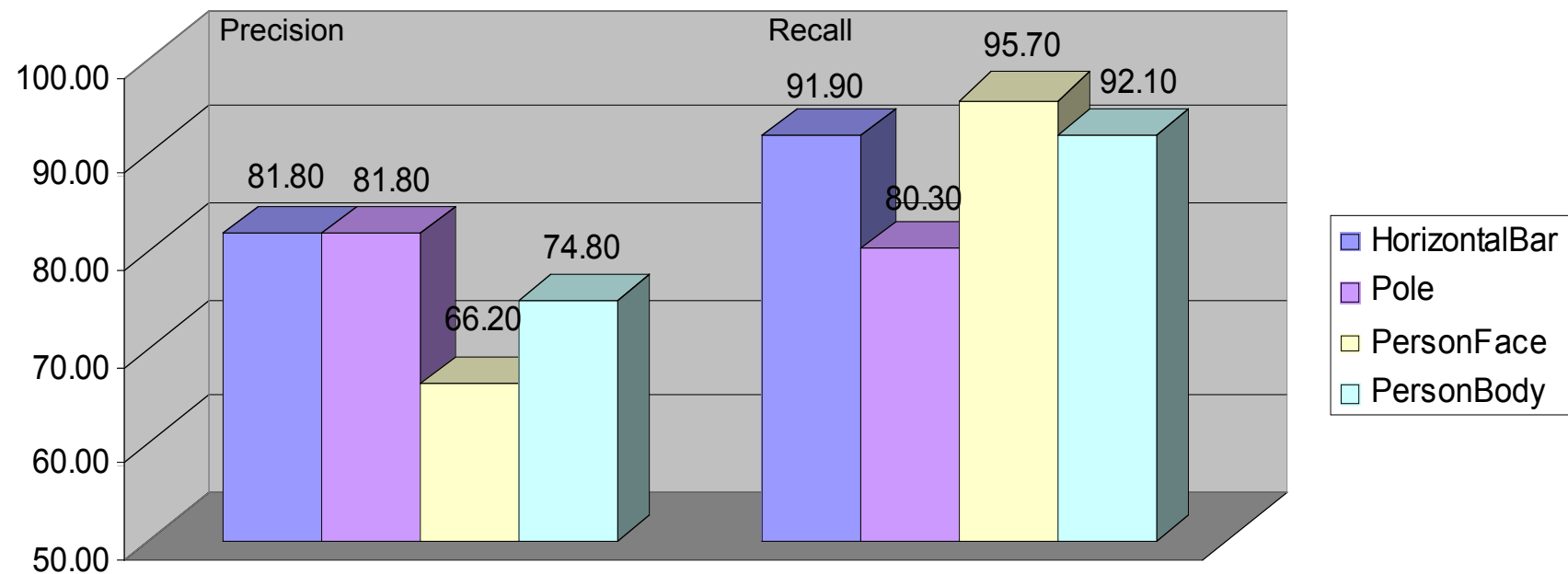
Information extraction from images

- Robust combination of region-based and holistic approaches
 - ♦ E.g. for body detection
 - foreground object detection through visual attention is combined with image segmentation and image segment classification, including merging of segments
- Variation of Hough transform for detection of elongated objects on integral projections of intensity images resulting from intensity derivatives



Evaluation of state of the art

Image extraction tool: Precision and recall at image level



$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

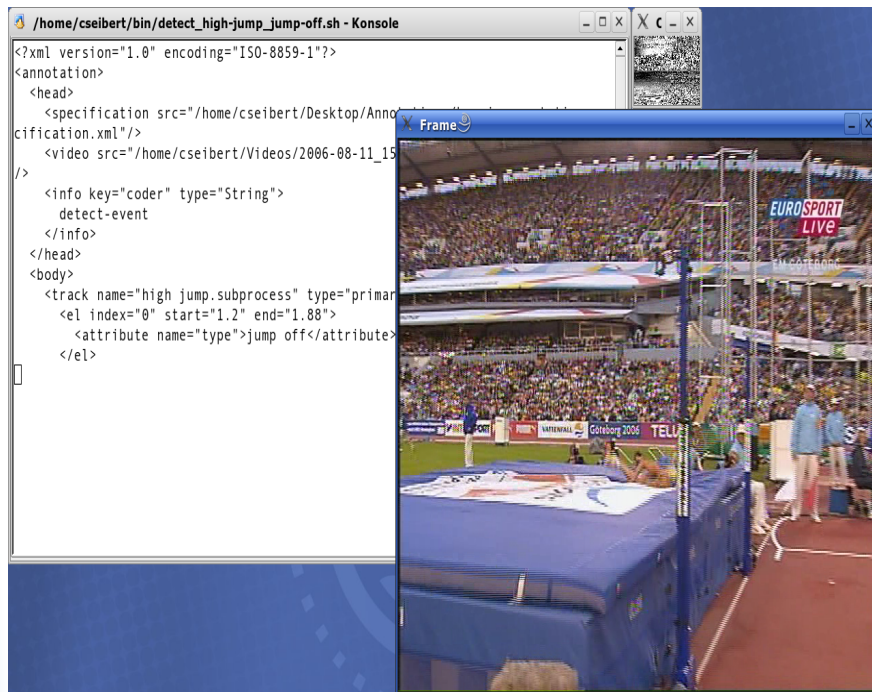
$$\text{fall-out} = \frac{|\{\text{irrelevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{irrelevant documents}\}|}$$

Extraction from video

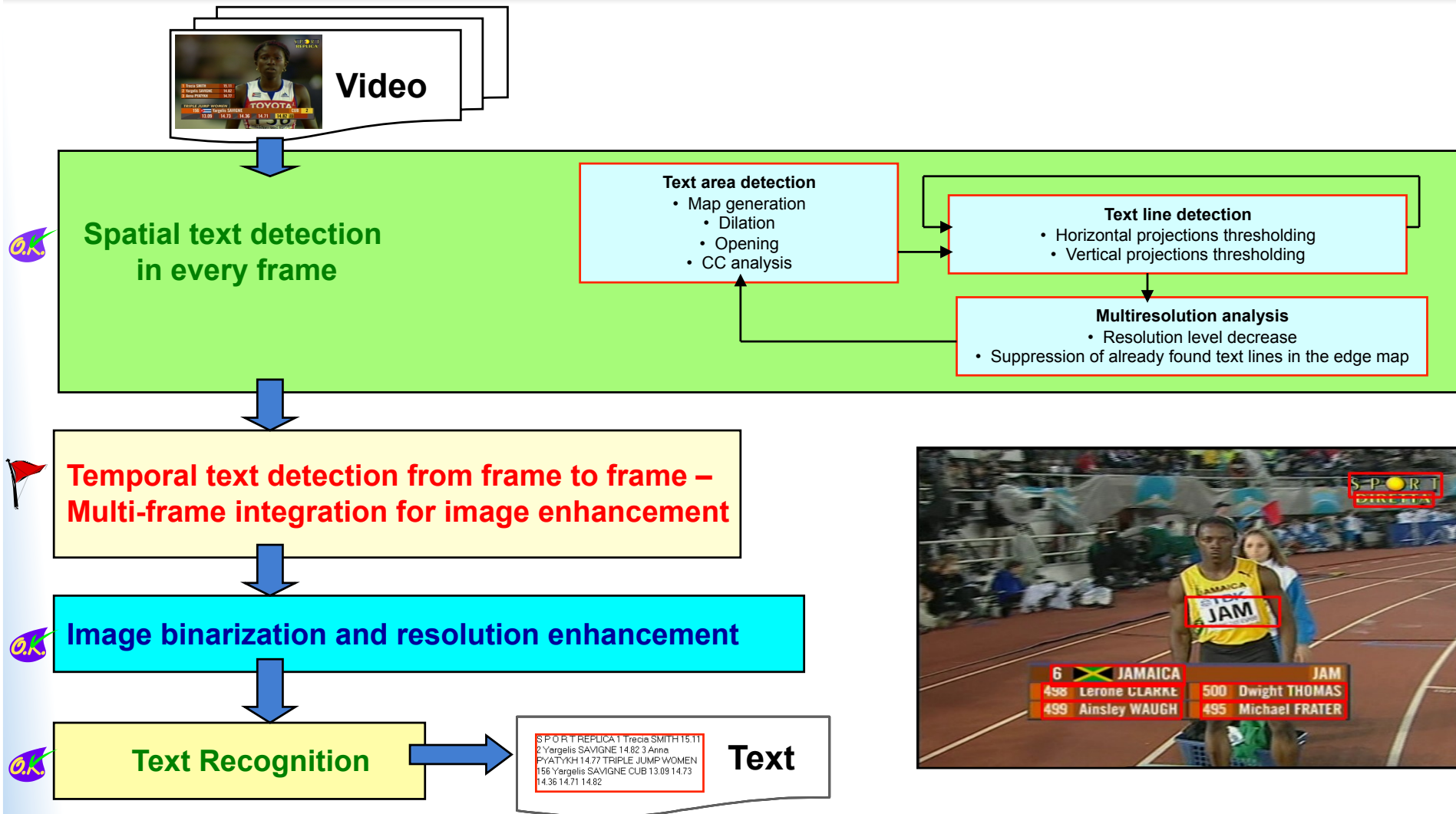
Aim:

- To identify MLCIs corresponding to phases of events, e.g. “swing and row” in pole vault or ‘jump off” in high jump
- To specify temporal relationships between these phases
- To track and identify objects in video scenes corresponding to MLCs and their spatio-temporal relations
- To detect areas in videos or images where artificial or scene text is present
- To extract the corresponding text

Video analysis tools

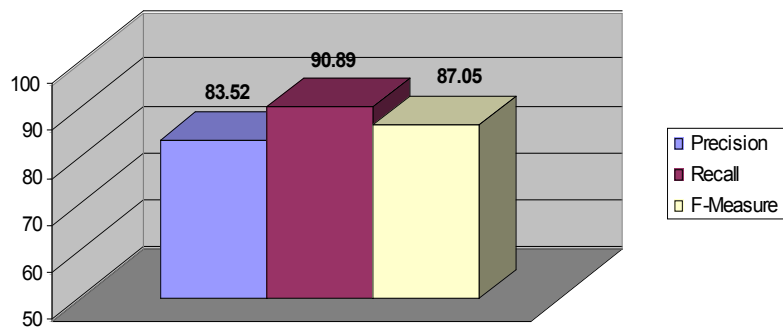


Video OCR tool: Functionality

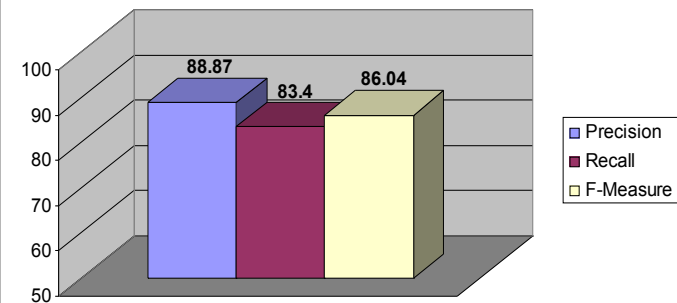


Video OCR tools: Evaluation

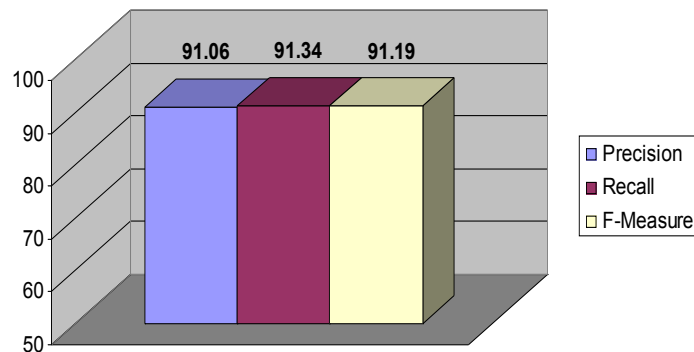
Video OCR tool: Text detection results, set 1: fonts of various sizes, presence of some scene text



Video OCR tool: Text detection results, set 2: larger fonts, presence of scene text



Video OCR tool: Text detection results, set 3: Artificial text, small fonts



F-measure

The weighted harmonic mean of precision and recall, the traditional F-measure or balanced F-score is:

$$F = 2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall}).$$

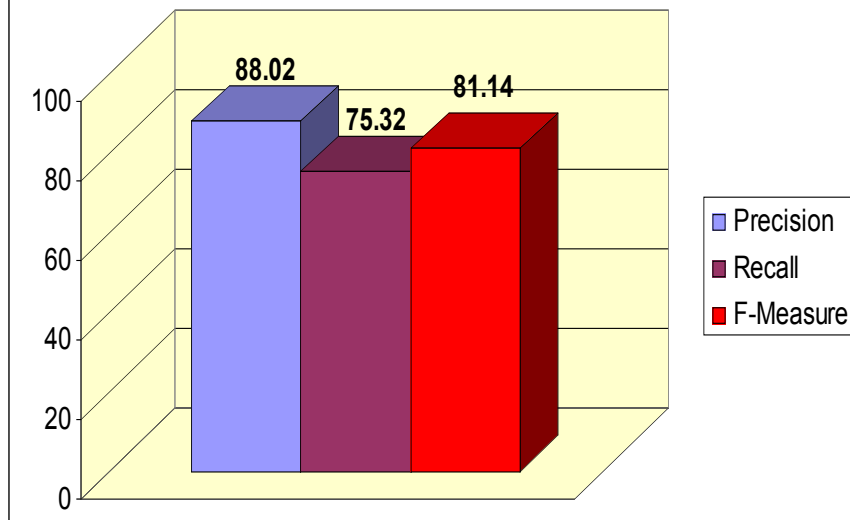
Extraction from text

- Aim:
 - ◆ Identification of MLCIs in textual documents (e.g. person names, sport names, event names etc.)
 - ◆ Identification of interesting relations concerning MLCIs (person name next-to sport name)

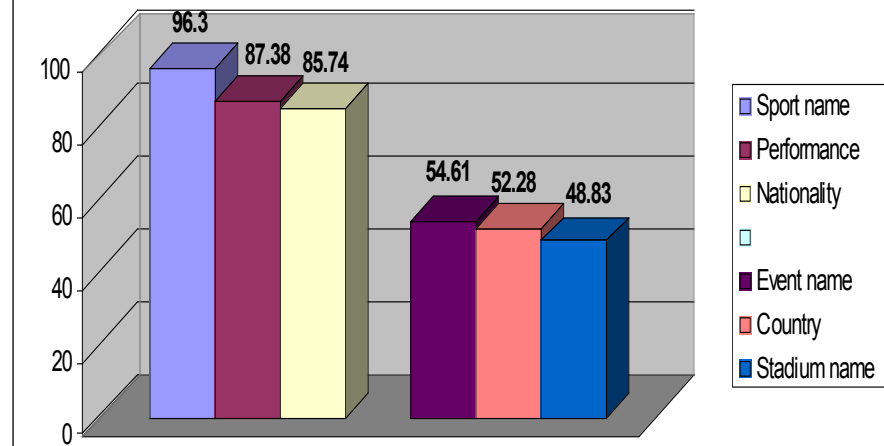


Text extraction: What to expect?

Text extraction tool: Performance

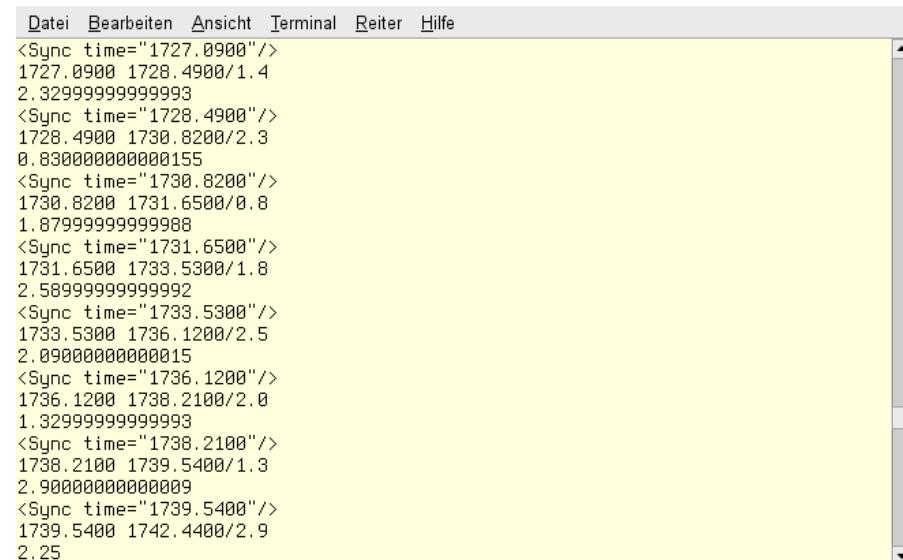


Text extraction tool: Best and worst detected categories (F-measures)



Extraction from audio

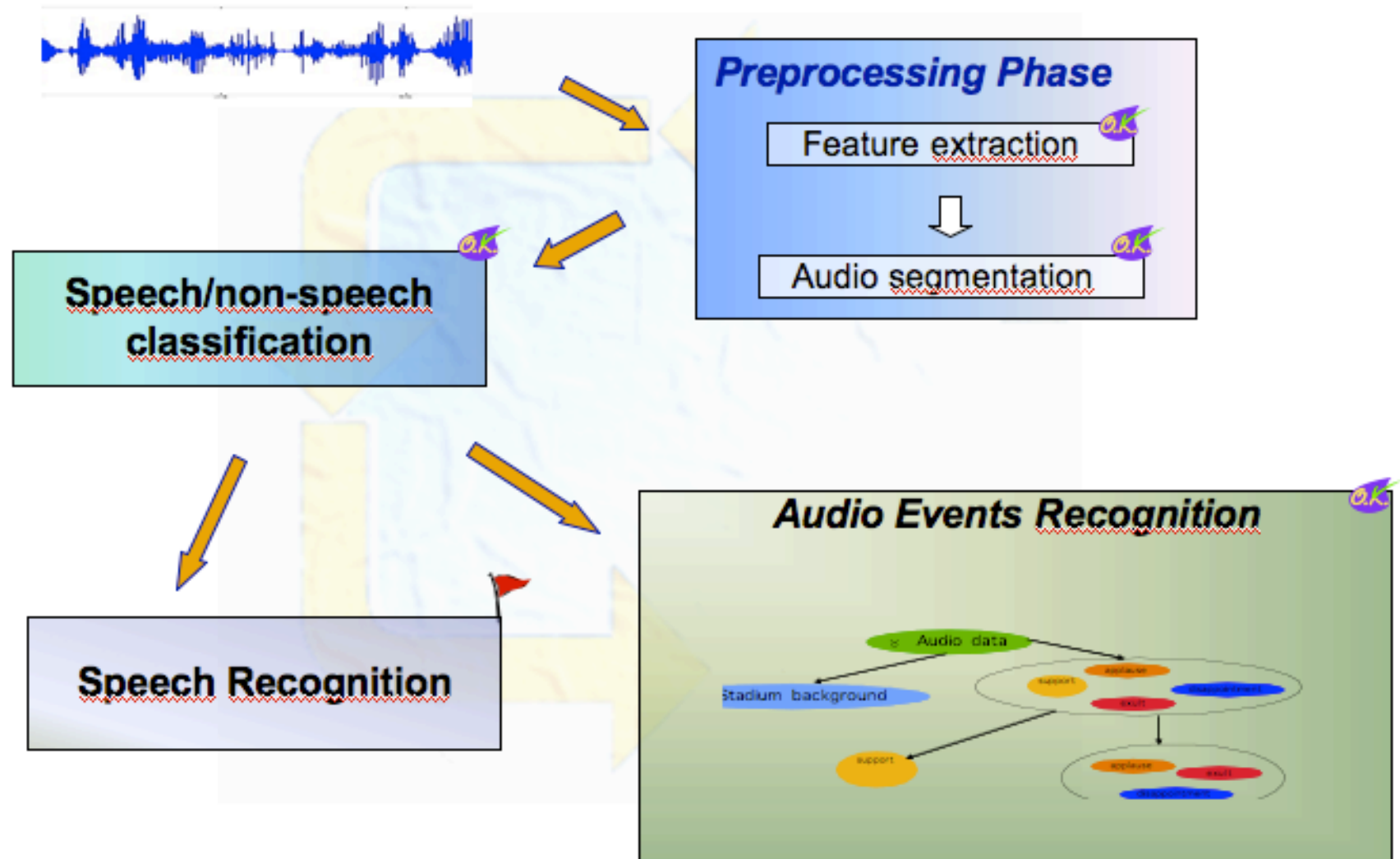
- Aim:
 - ♦ To discriminate between speech and non-speech segments of audio transcripts
 - ♦ To detect MLCIs as events in the non-speech part of the stream, such as Applause, Support, Stadium background audio and their temporal relations
 - ♦ To identify MLCIs in the speech transcript, e.g. names of athletes, performances etc



The screenshot shows a window with a menu bar (Datei, Bearbeiten, Ansicht, Terminal, Reiter, Hilfe) and a text area containing a list of synchronization events. Each event is represented by a line of text starting with a sync tag, followed by a timestamp, a range, and a value.

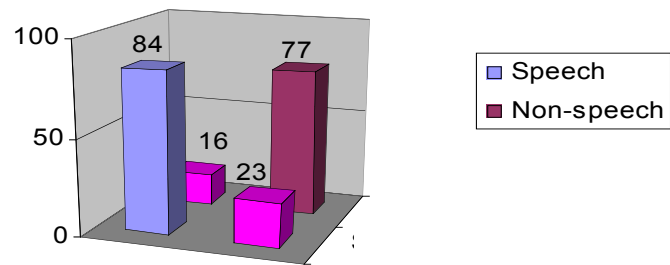
```
<Sync time="1727.0900"/>  
1727.0900 1728.4900/1.4  
2.329999999999993  
<Sync time="1728.4900"/>  
1728.4900 1730.8200/2.3  
0.83000000000000155  
<Sync time="1730.8200"/>  
1730.8200 1731.6500/0.8  
1.879999999999998  
<Sync time="1731.6500"/>  
1731.6500 1733.5300/1.8  
2.589999999999992  
<Sync time="1733.5300"/>  
1733.5300 1736.1200/2.5  
2.0900000000000015  
<Sync time="1736.1200"/>  
1736.1200 1738.2100/2.0  
1.329999999999993  
<Sync time="1738.2100"/>  
1738.2100 1739.5400/1.3  
2.900000000000009  
<Sync time="1739.5400"/>  
1739.5400 1742.4400/2.9  
2.25
```

Audio analysis tool: Functionality

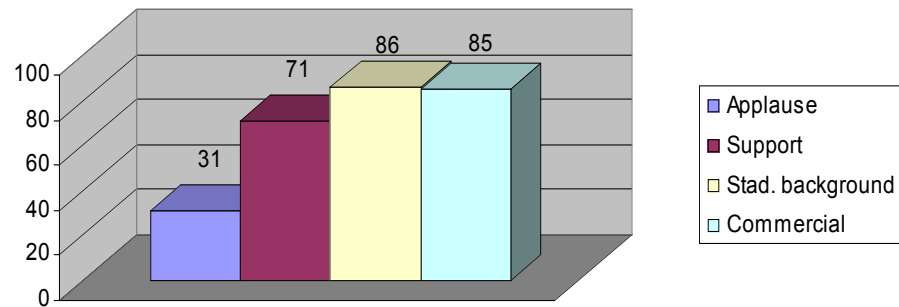


Audio analysis tool: Evaluation

Confusion matrix for speech-non speech discrimination



F-measures for classification of non-speech events



Confusion Matrix

In the example confusion matrix below, of the 8 actual cats, the system predicted that three were dogs, and of the six dogs, it predicted that one was a rabbit and two were cats. We can see from the matrix that the system in question has trouble distinguishing between cats and dogs, but can make the distinction between rabbits and other types of animals pretty well.

Example confusion
matrix

	Cat	Dog	Rabbit
Cat	5	3	0
Dog	2	3	1
Rabbit	0	2	11

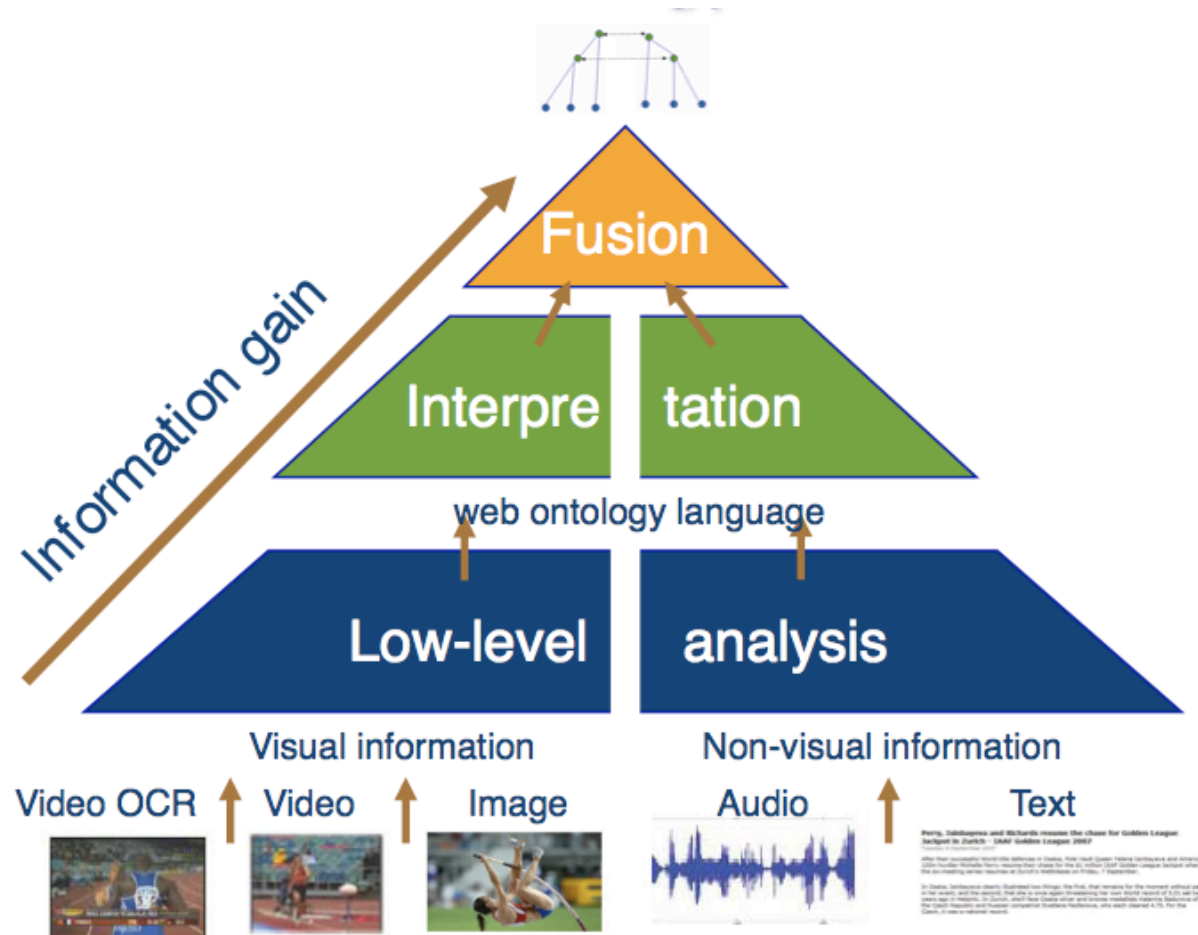
- CF matrix is a visualization tool typically used in supervised learning

Summary

- Analysis techniques for different modalities
 - ♦ Still images
 - ♦ Videos
 - ♦ Audio
 - ♦ Text
- Low-level features quite reliably detected
 - ♦ Analysis could still be improved
 - ♦ From bias to considering context information
- But: High-level interpretation missing
 - ♦ Formalization required
 - ♦ Semantic gap
 - ♦ Feedback between interpretation and analysis
- Fusion of analysis and interpretation results for multiple modalities still an open issue

Information Extraction (IE)

Different levels of ontology based IE:

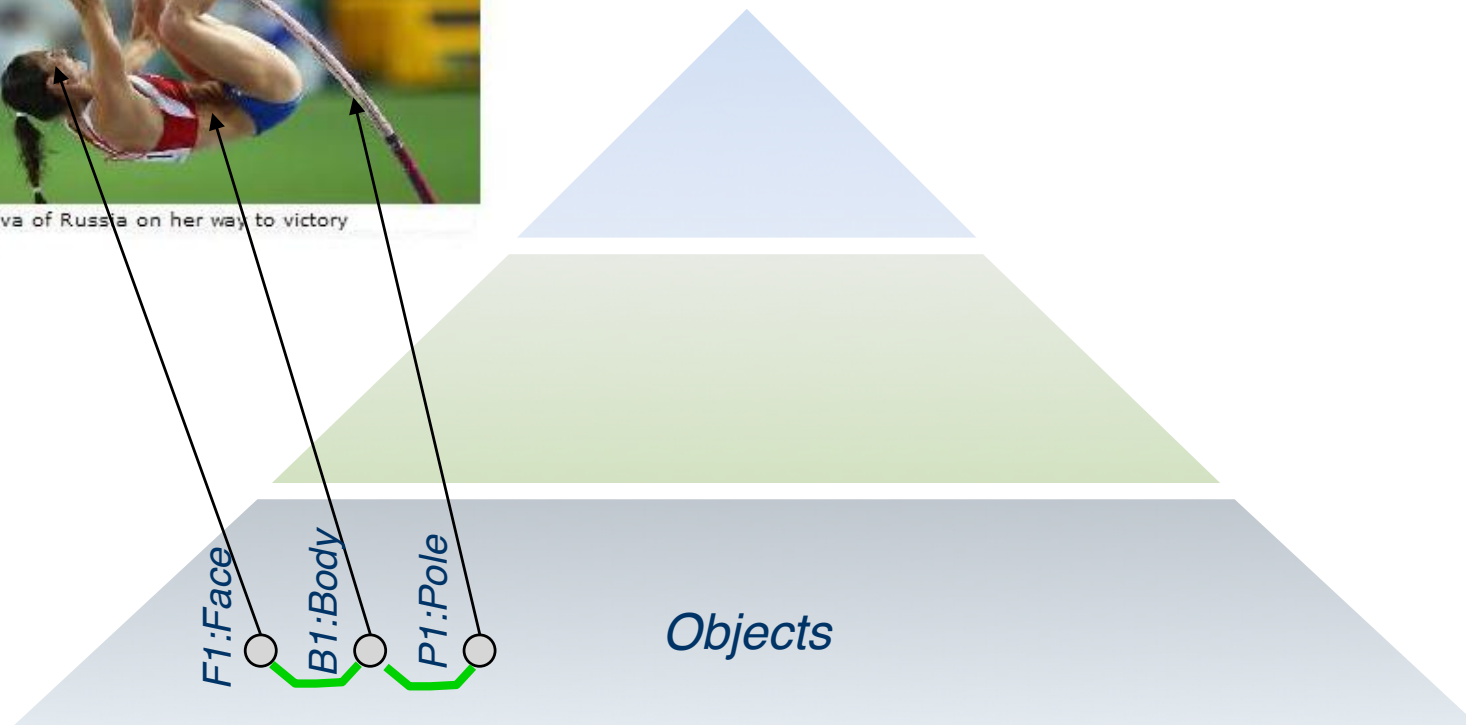


Observable objects

- Low-level analysis of visual modalities



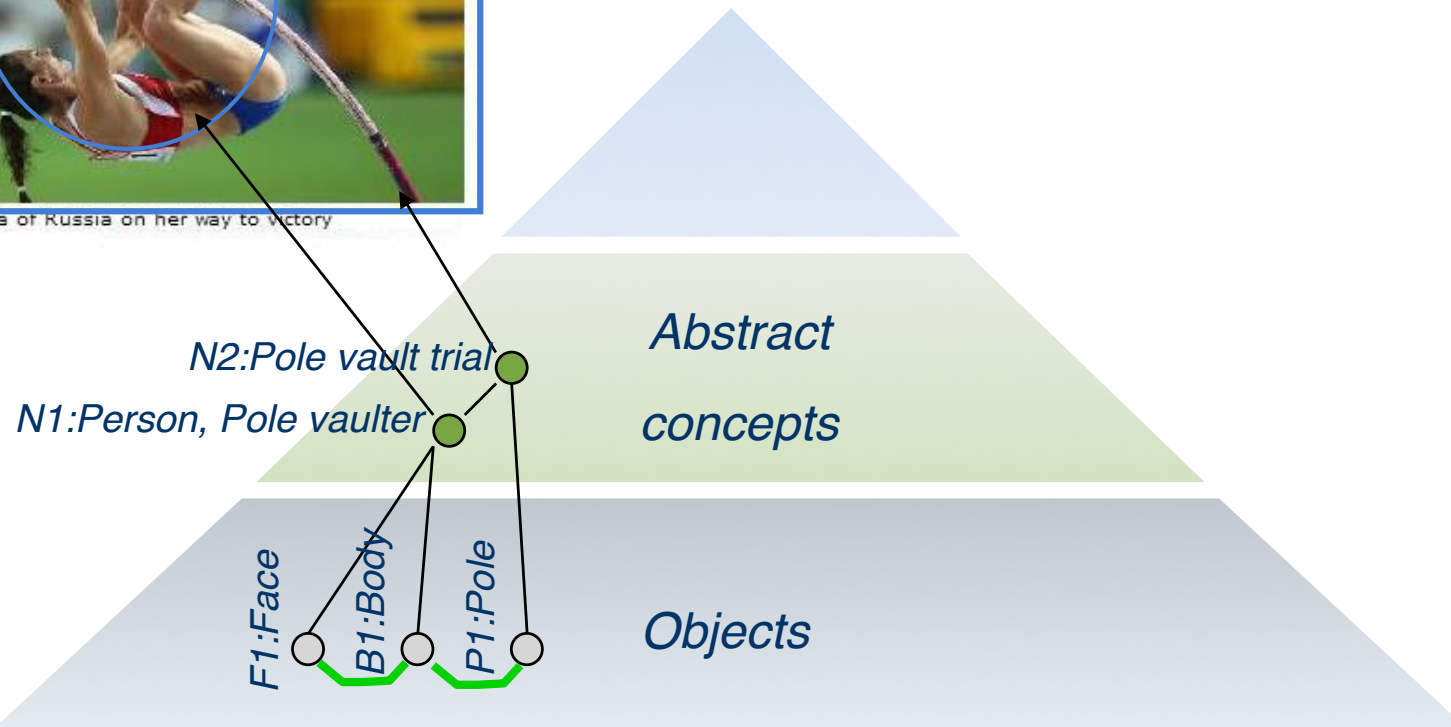
Yelena Isinbayeva of Russia on her way to victory
(Getty Images)



Interpretation = Explanation



Yelena Isinbayeva of Russia on her way to victory
(Getty Images)



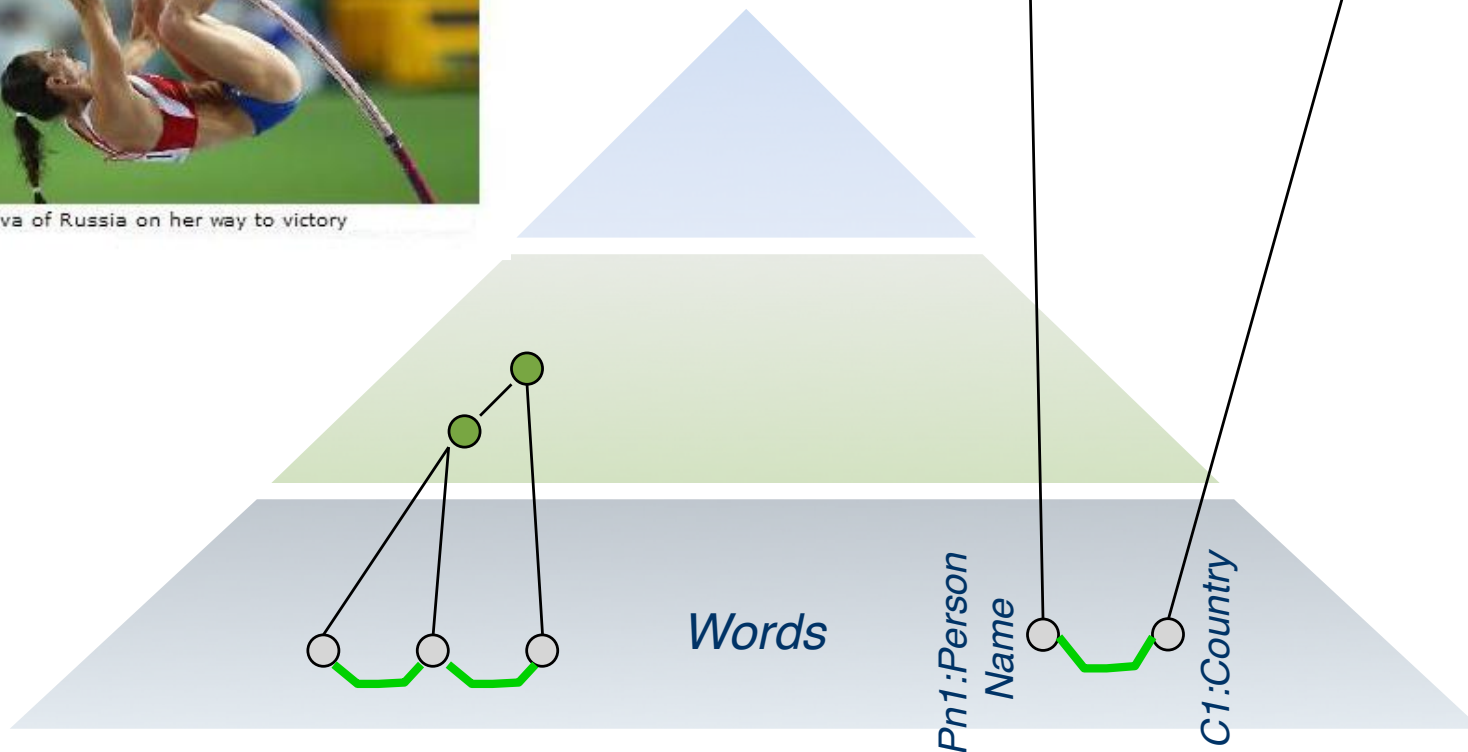
Text modality

- Low-level analysis of text



Yelena Isinbayeva of Russia on her way to victory
(Getty Images)

Yelena Isinbayeva of Russia on
her way to victory (Getty Images)



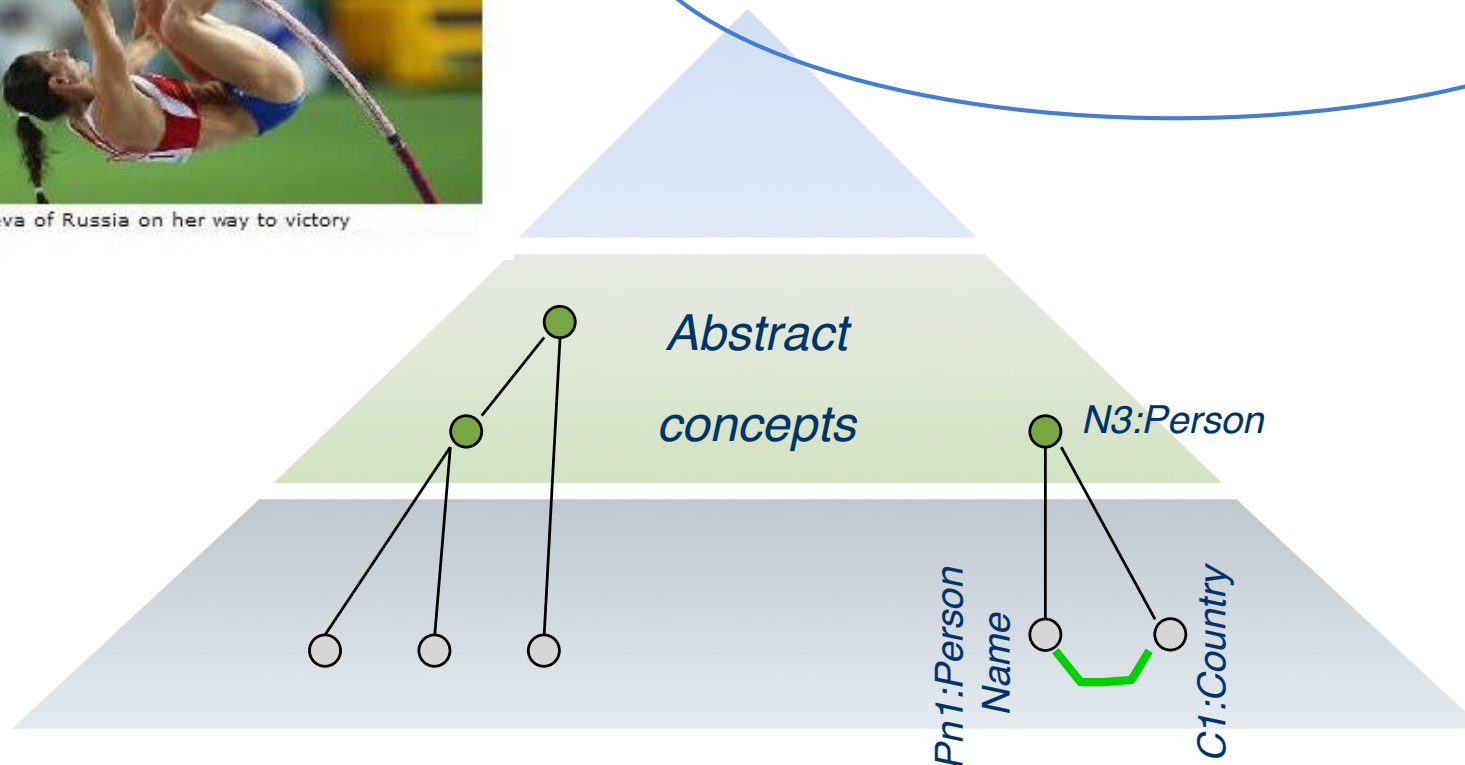
Text modality (2)

- Text interpretation



Yelena Isinbayeva of Russia on her way to victory
(Getty Images)

Yelena Isinbayeva of Russia on
her way to victory (Getty Images)

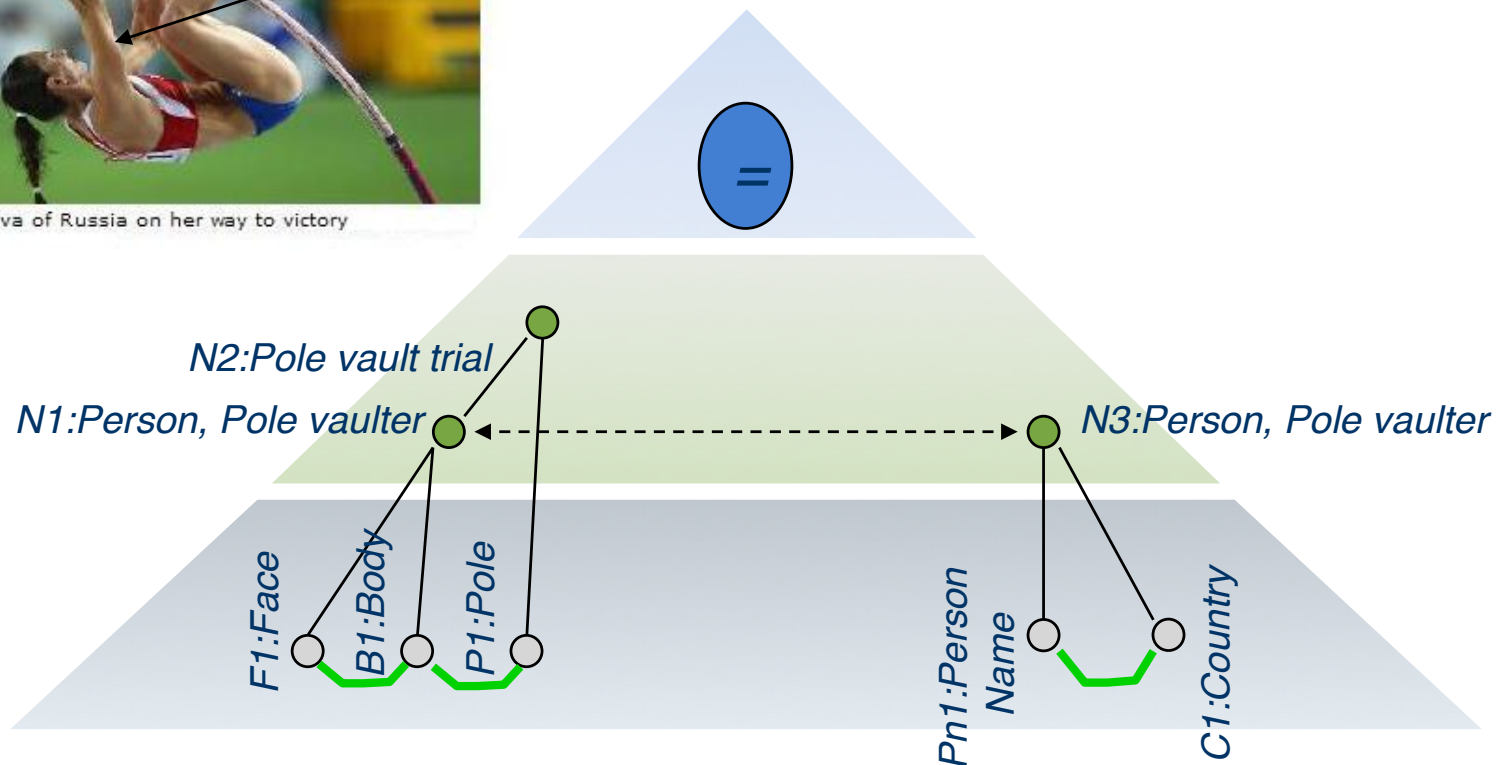


Fusion

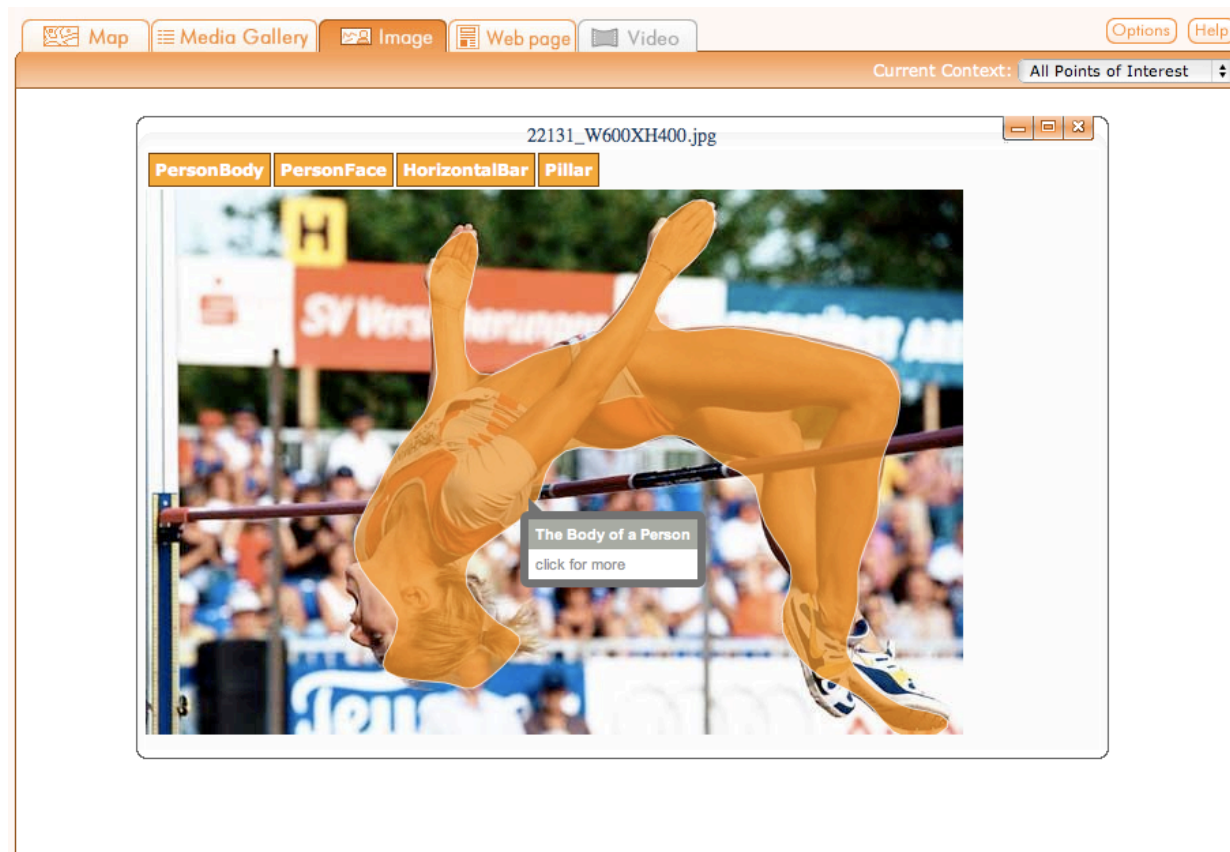


Yelena Isinbayeva of Russia on her way to victory
(Getty Images)

Yelena Isinbayeva of Russia on
her way to victory (Getty Images)



Using annotations: BSB-Demo



Dynamic suggestion of related information

- Exploits explicit and implicit information to provide for:
 - ✓ context related advertisement
 - ✓ to suggest related information.

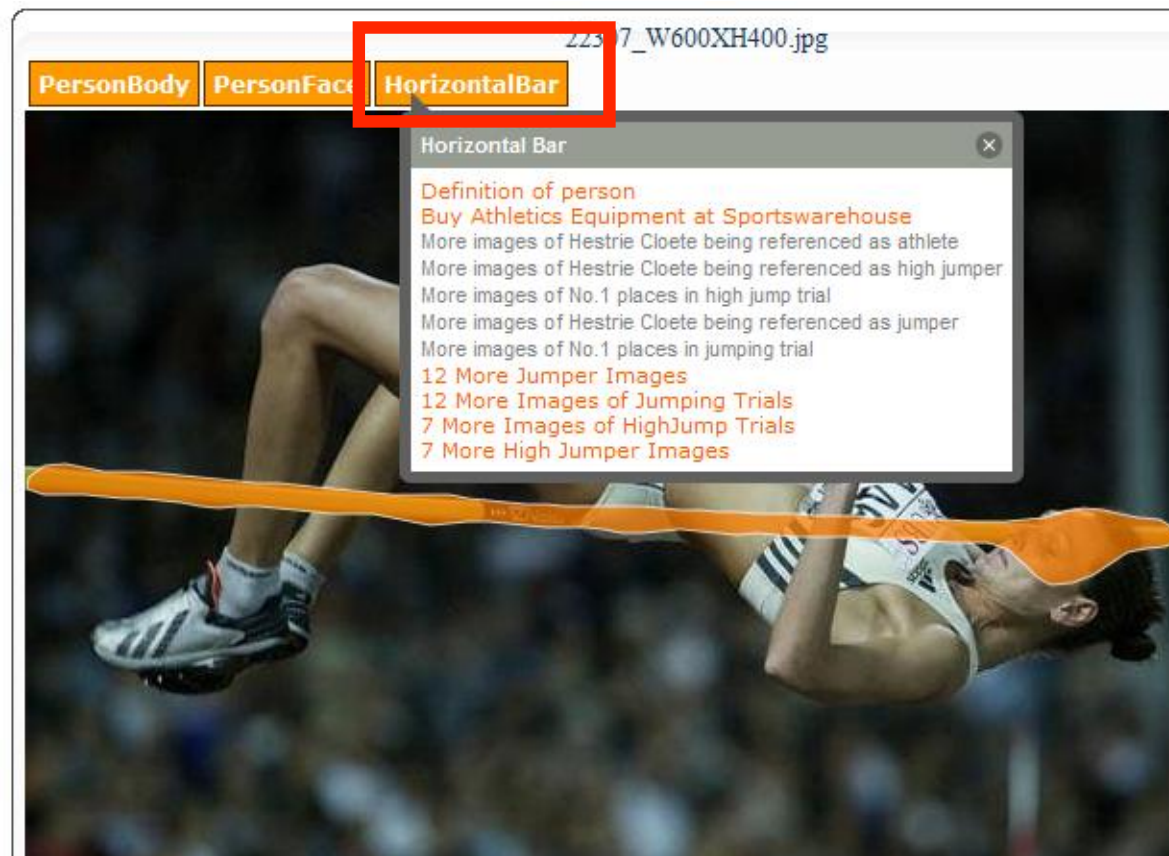
The screenshot shows a web browser window with the address bar displaying `http://repository.boemie.org:8..bad-a963-91de11a4825e.e-1.html`. The main content area features a news article titled "Mutola stays on track for US\$ 1 Million Jackpot". The article text includes mentions of "women's 800m", "Maria Mutola", "gold", "US\$ 1 Million Golden League Jackpot", "Hestrie Cloete", "South Africa", "2.03 to win in Zurich", "Mozambique", "Mutola - first 1:59.93", "Ster Graf - second 2:00.52", "third Claudia Gesell - 2:01.03", "El Guerrouj", "Hicham El Guerrouj", "men's 1500m in London (8 August)", "Moroccan's 3:29.13", "winning", "Paris", "5000m", "Bernard Lagat", and "Kenya". There are two small images: one of a high jumper and one of a runner. A "Ranking" popup window is open over the article, displaying the following content:

Ranking

- Definition of an athlete
- More articles about high jump trials
- More references to 1 places
- 1 More articles about jumping trials
- 6 More articles about other high jump athletes
- 7 More articles about other jump athletes

Dynamite

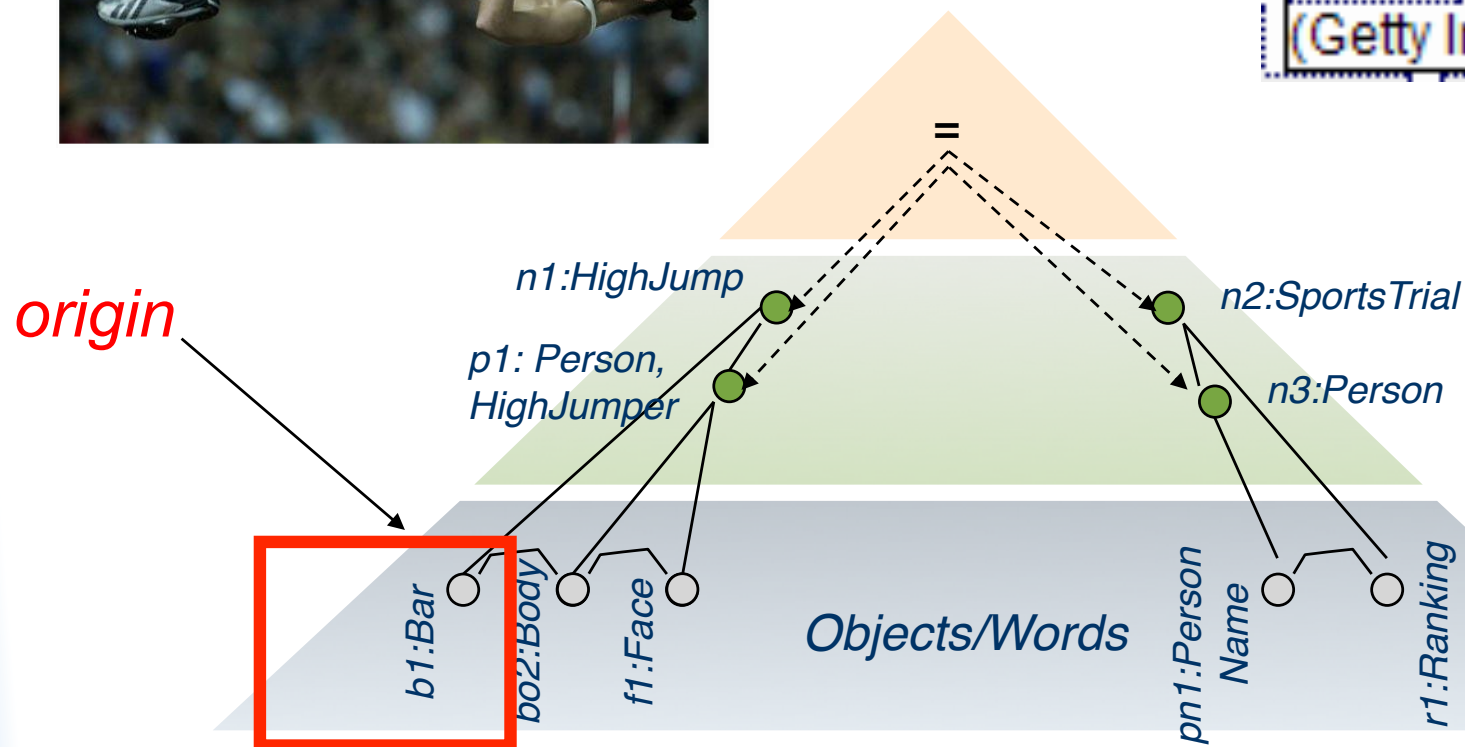
- Determines applicable services to activate in a context menu.



Interpretation context determines applicable services



Hestrie Cloete of South Africa clears 2.03 to win in Zurich (Getty Images)



Service definitions (1)



ServiceId: 1

Menu-name: Buy Athletics Equipment at Sportswarehouse

Arguments: x : AthleticsEquipment

Type: WebNavigation

URL: <http://www.sportswarehouse.co.uk/acatalog/Athletics>

*Determines
applicability*

Service definitions (2)

“More images of Hastrie Cloete being referenced as high jumper”

ServiceId: 2

Aux: %x% = getFiller (x, aeo:hasPersonNameValue)

Menu-name: More images of %x% being referenced as high jumper

Arguments: x : PersonName, y : HighJumper

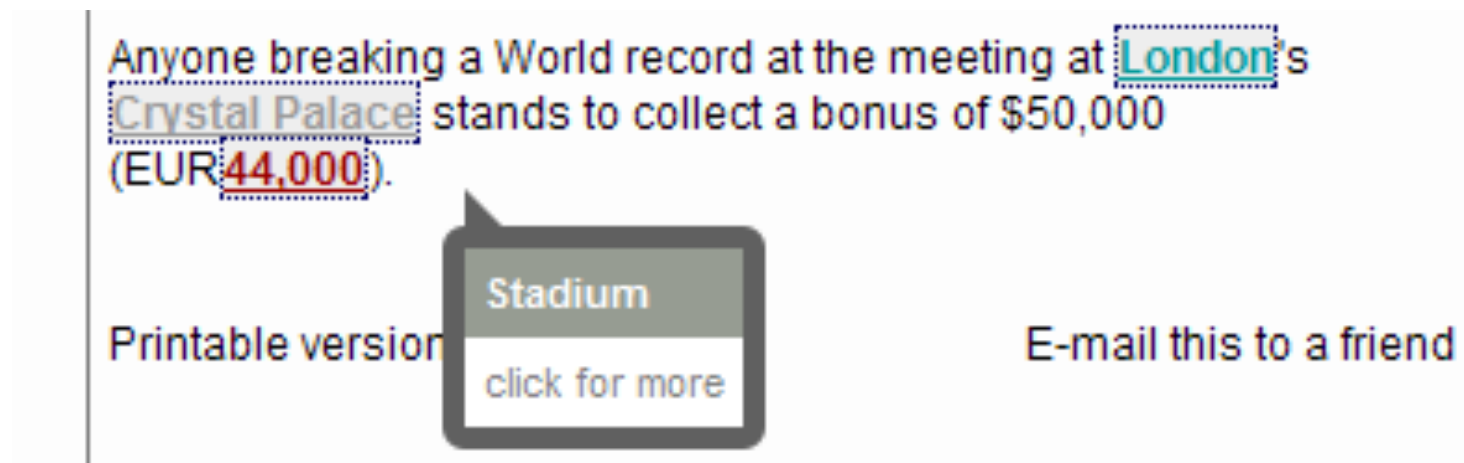
Type: RepositoryNavigation

Query:

```
SELECT DISTINCT ?u WHERE {  
    ?w rdf:type mco:Image .  
    ?w mco:hasURL ?u .  
    ?w mco:depicts ?y .  
    ?y rdf:type aeo:HighJump .  
    ?y aeo:hasParticipant ?z .  
    ?z aeo:hasPersonsName ?pn .  
    ?pn aeo:hasPersonsNameValue %x%. }
```

Geo-localization of media

- Geographic references extracted from non-visual content, e.g., city, country, point of interest.
- Geography-aware information navigation / retrieval
- Usage of TeleAtlas GIS services to obtain coordinates for a geographic reference.
- Example: Web pages



Geography-aware IR

- Demonstrates that semantic tagging of map information can be extended to cover multiple types of media, e.g. video, image, text.
 - ✓ Without the need for online communities

