Bootstrapping Ontology Evolution with Multimedia Information Extraction

boemie

Bootstrapping Ontology Evolution with Multimedia Information Extraction

Information Society Technologies

The facts

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

- STREP, IST-2004-2.4.7 “Semantic-based Knowledge and Content Systems”
- Start: March 1, 2006
- End: February 28, 2009
- Budget: € 5.075.678
- EU Funding: € 3.150.000
- More than 40 people active in the project
- Project portal: http://www.boemie.org/

Information Society Technologies

Consortium

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

- Inst. of Informatics & Telecommunications, NCSR "Demokritos" (SKEL & CIL), Greece (Coordinator)
- Fraunhofer Institute for Media Communication (NetMedia), Germany
- Dip. di Informatica e Comunicazione, University of Milano (ISLab), Italy
- Inst. of Telematics and Informatics CERTH (IPL), Greece
- Hamburg University of Technology (STS), Germany
- Tele Atlas SA, The Netherlands

Information Society Technologies

Vision

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

- Pave the way towards automation of the knowledge acquisition process from multimedia content.
- Break new ground by introducing and implementing the concept of evolving ontologies for multimedia.
- Make domain-specific semantic webs feasible with limited human effort.

Information Society Technologies

Objectives

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

- Providing technology to represent and evolve domain-specific ontologies for multimedia.
- Moving from low-level, general-purpose, single-modality feature extraction towards semantic, multimedia analysis.
- Robust and scalable ontology-driven multimedia content extraction through ontology evolution.

Information Society Technologies

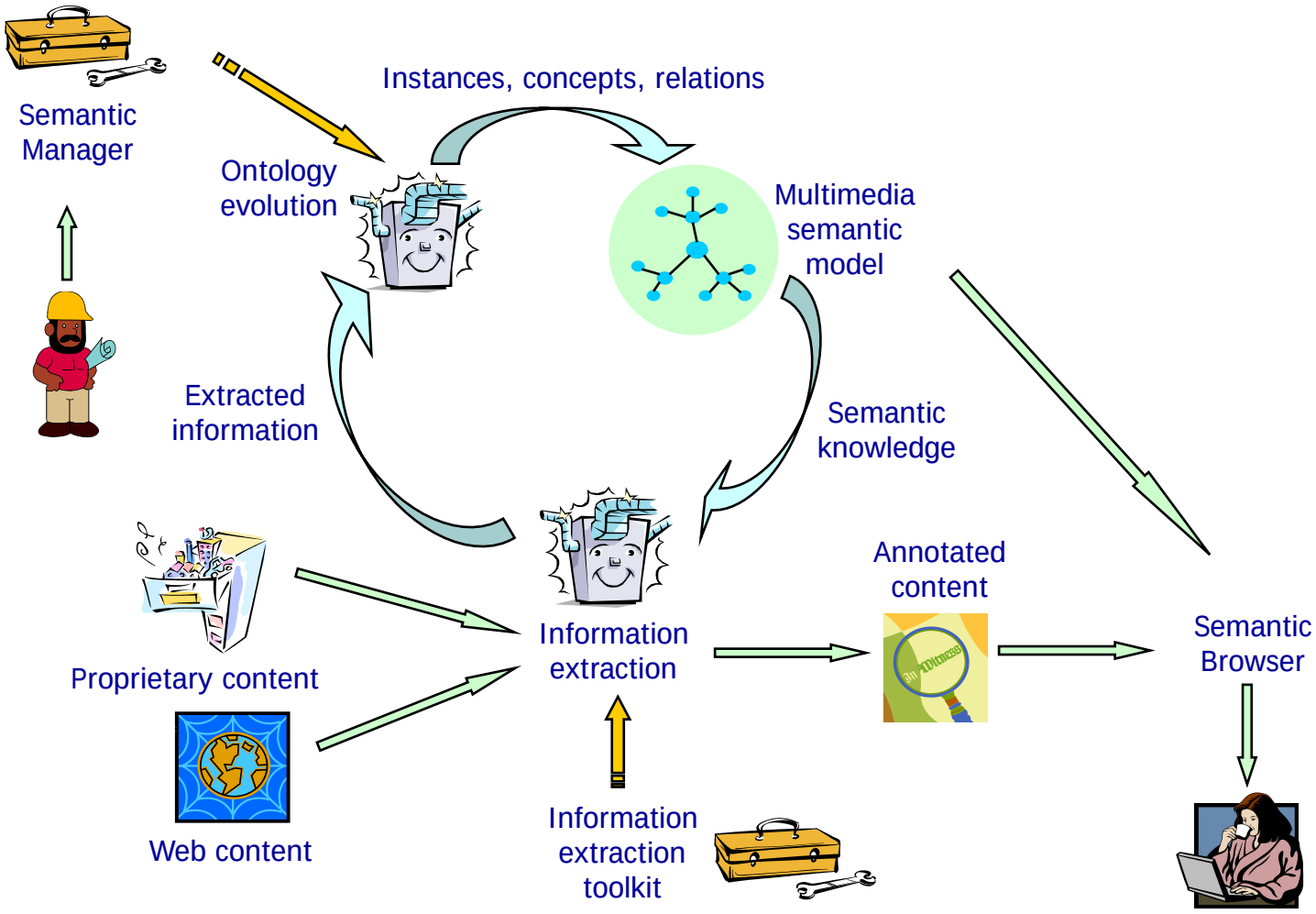
Approach

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

- Driven by domain-specific ontologies, BOEMIE information extraction systems aim to identify high-level semantic features in image, video, audio and text and fuse these features for optimal extraction.
- The ontologies are continuously populated and enriched using the extracted semantic content.
- This is a bootstrapping process, since the enriched ontologies are in turn used to drive the multimedia information extraction system.

Information Society Technologies

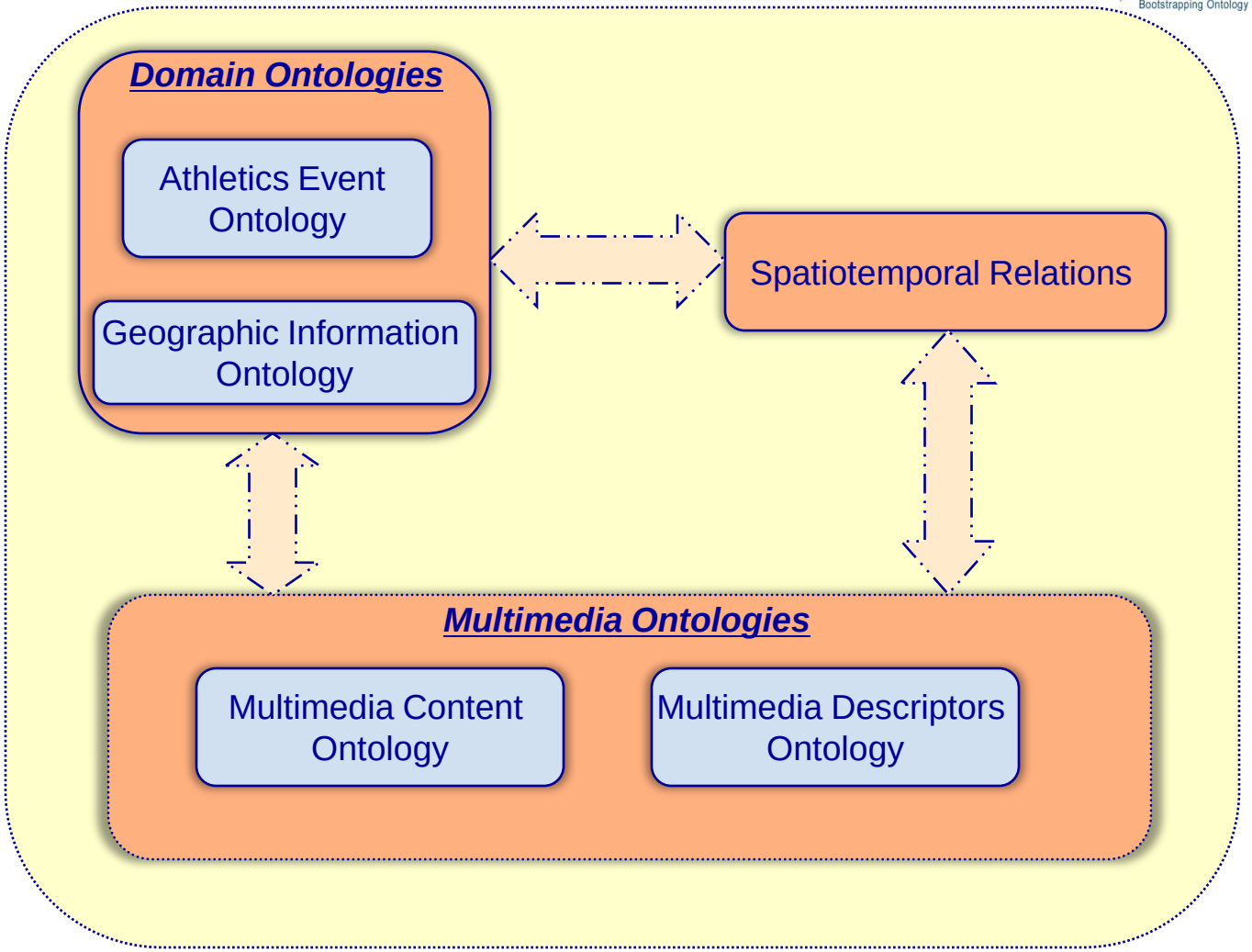
Approach


Semantic Manager
Instances, concepts, relations
Ontology evolution
Multimedia semantic model
Extracted information
Semantic knowledge
Annotated content
Proprietary content
Information extraction
Semantic Browser
Web content
Information extraction toolkit

Multimedia semantic model

- A geographic ontology, e.g. about landmarks.
- An event ontology, e.g. about athletics events.
- Multimedia content and descriptor ontologies
- Potential contribution:
  - Uncertainty in concept descriptions.
  - Spatial and temporal relations.
  - Efficient (leightweight) representation.

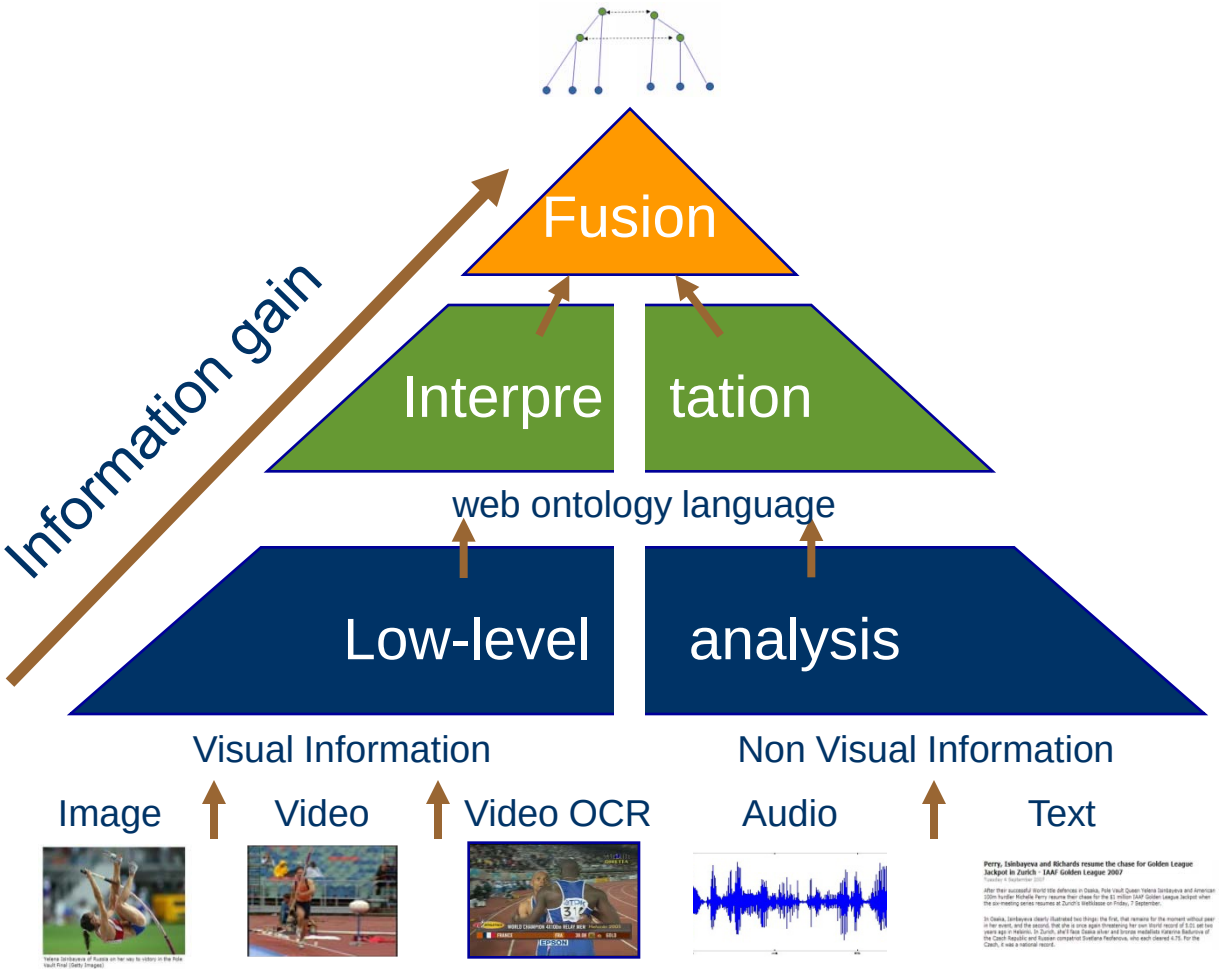
Multimedia semantic model



Semantics extraction

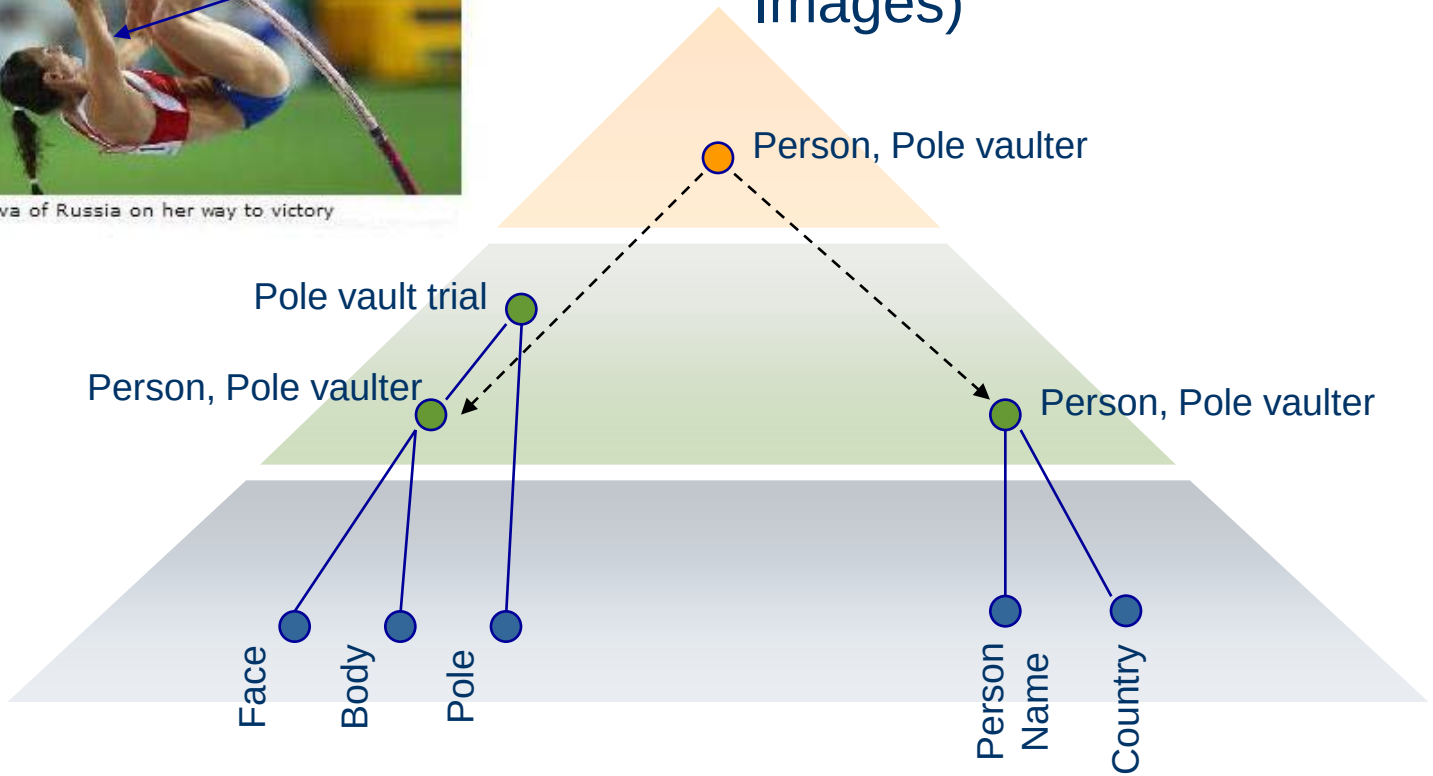
- Emphasis on fusion of multiple modalities, using reasoning and handling uncertainty.
- Contribution to the state of the art in visual content analysis, due to its richness and the difficulty of extracting semantics.
- Non-visual enriches semantic annotation through fusion.
- Two dominating content formats: Web pages with images and video (mainly proprietary).

Semantics extraction



Semantics extraction

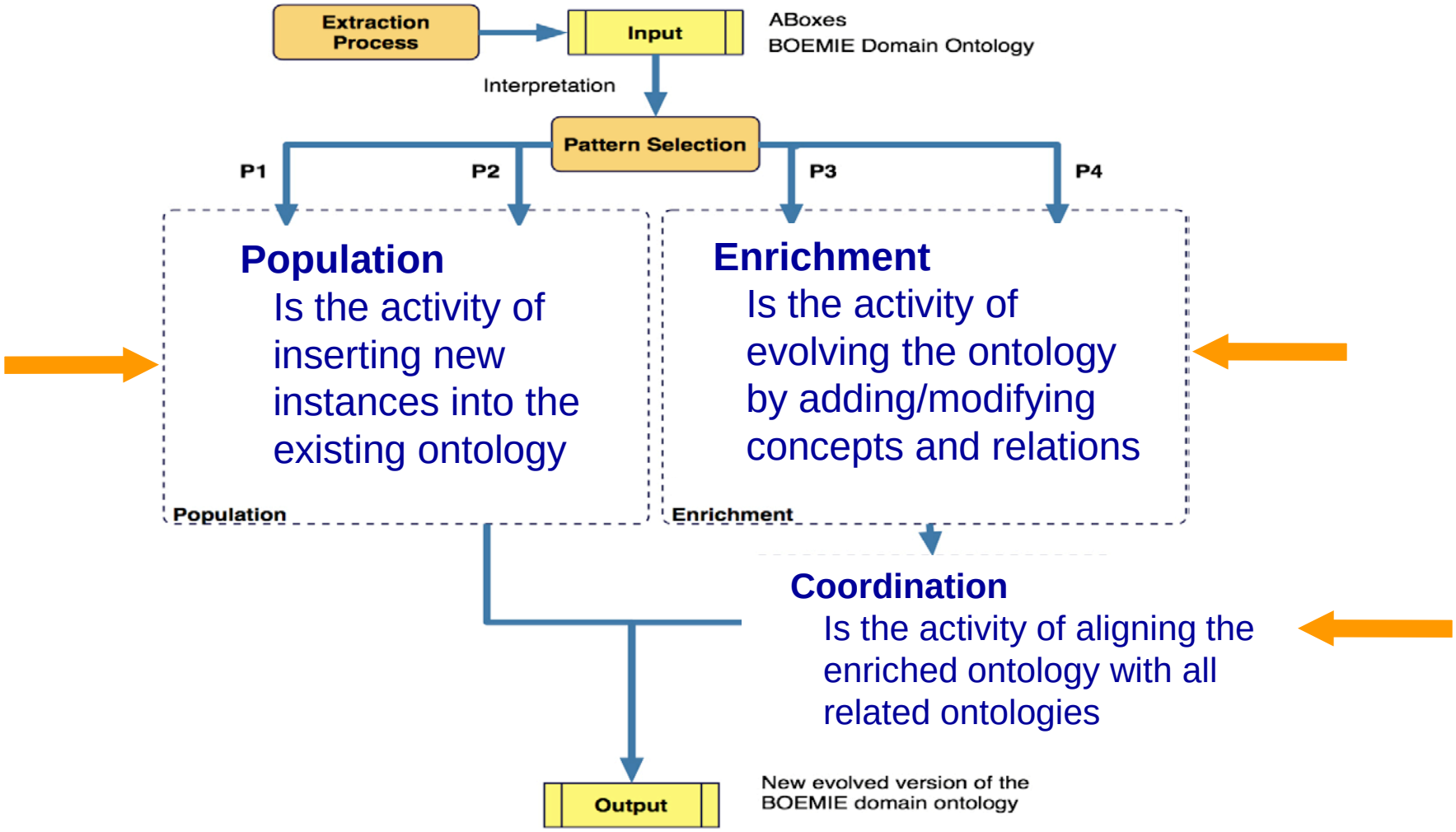
Yelena Isinbayeva of Russia on her way to victory (Getty Images)



Ontology evolution

- Ontology population from multimedia content.
- Combination of reasoning and statistics for enrichment and coordination.
- Matching and coordination support for population and enrichment.

Ontology evolution



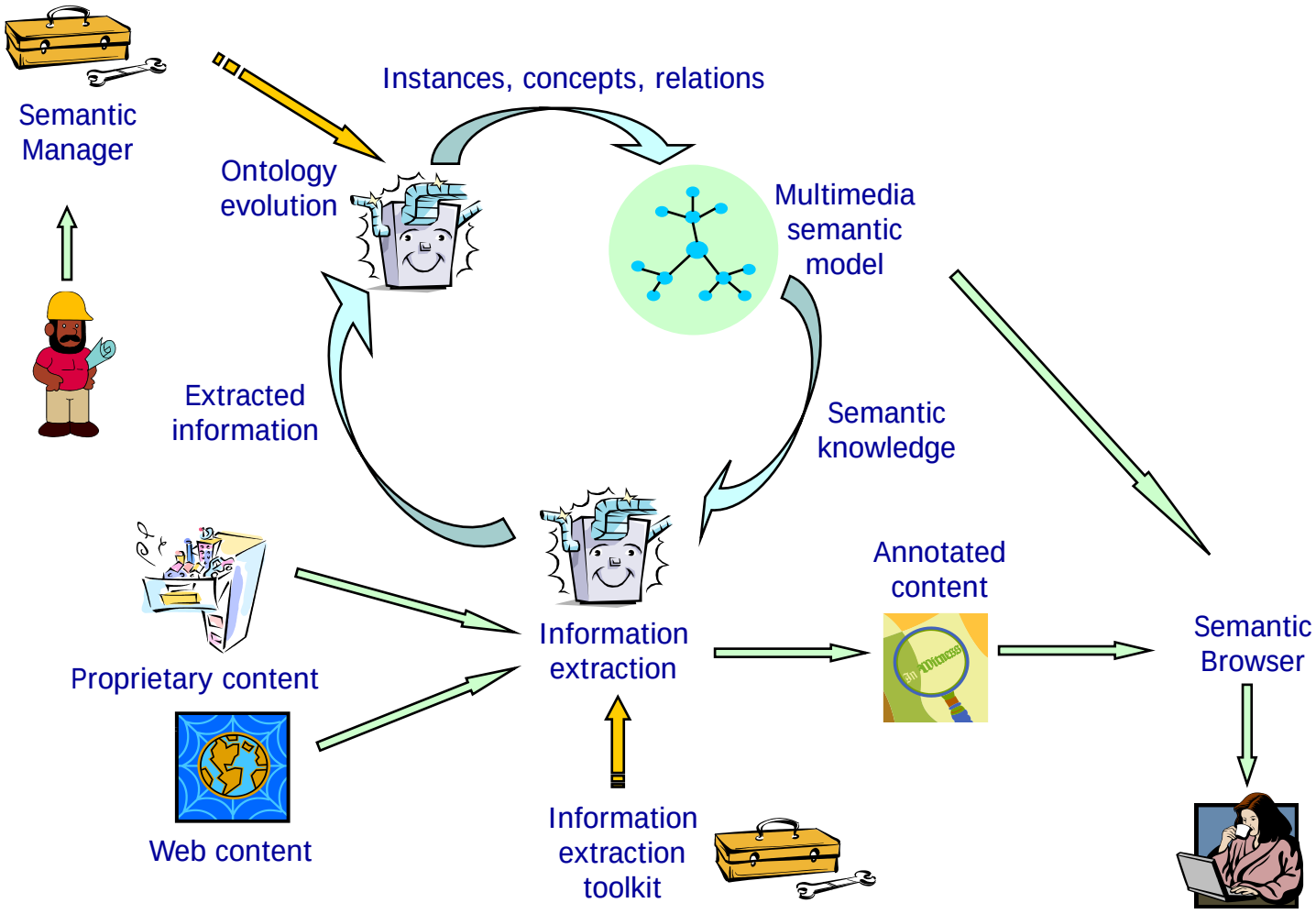
System integration

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

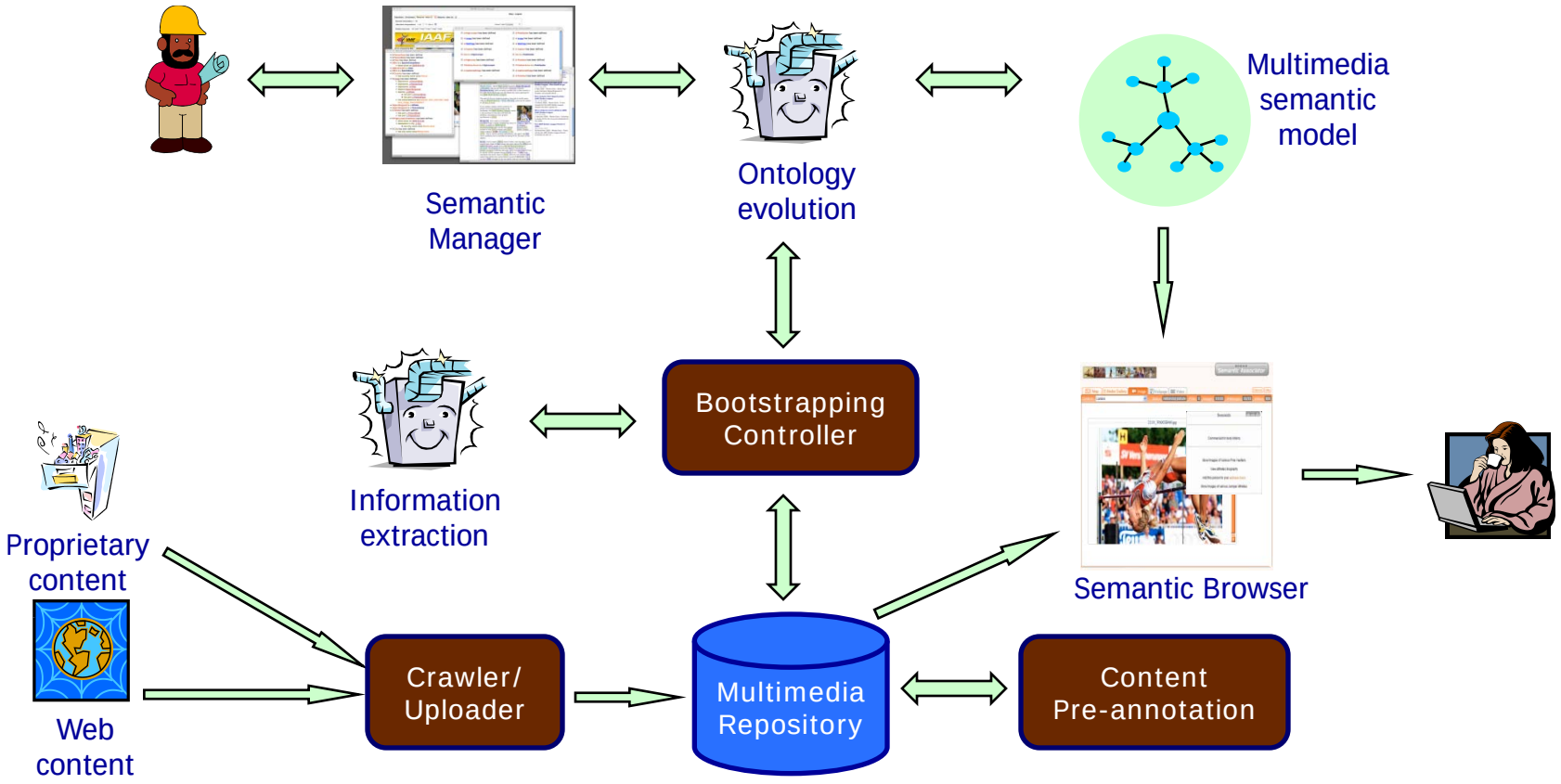
- Implementation of the bootstrapping process, integrating semantic extraction and ontology evolution, through the semantic model.
- Crawling for content collection and content quality assessment.
- Demonstration of added value for the content provider (semantic content management) and the end user (semantic content browsing)!

Information Society Technologies

System integration



System integration



Achievements

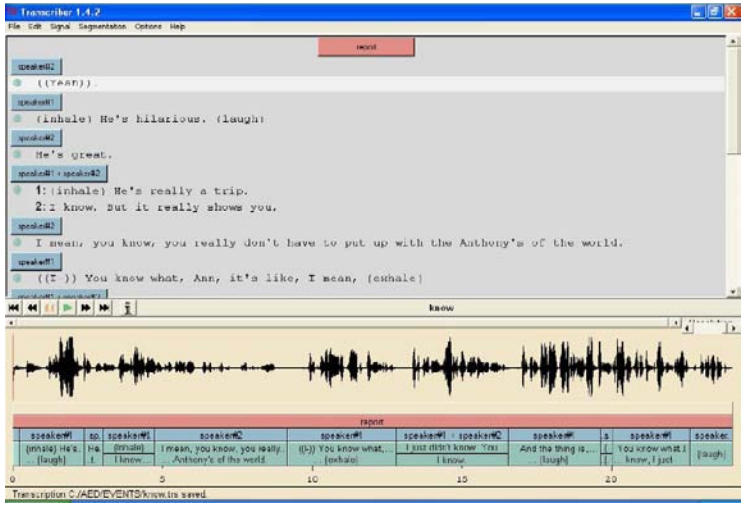
boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

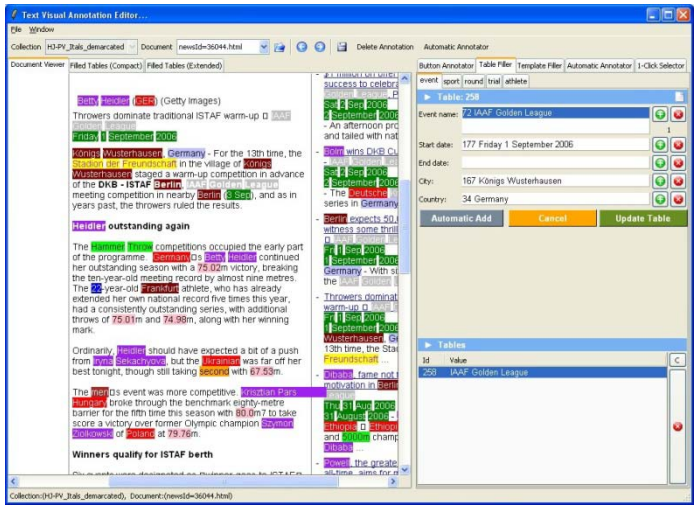
- Multimedia, geographic and athletics ontologies, integrated into a common BOEMIE semantic model.
- Novel extraction methods for images, video, audio, video OCR, integrated into the extraction toolkit.
- Advanced reasoning techniques for multimedia interpretation, based on abductive methods.
- Innovative ontology population, enrichment and coordination techniques, integrated into the ontology evolution toolkit.
- Advanced content management system, implementing the bootstrapping process.
- Prototype to be evaluated by external users.
- User-friendly annotation tools for each modality.
- Large annotated corpus for various modalities.

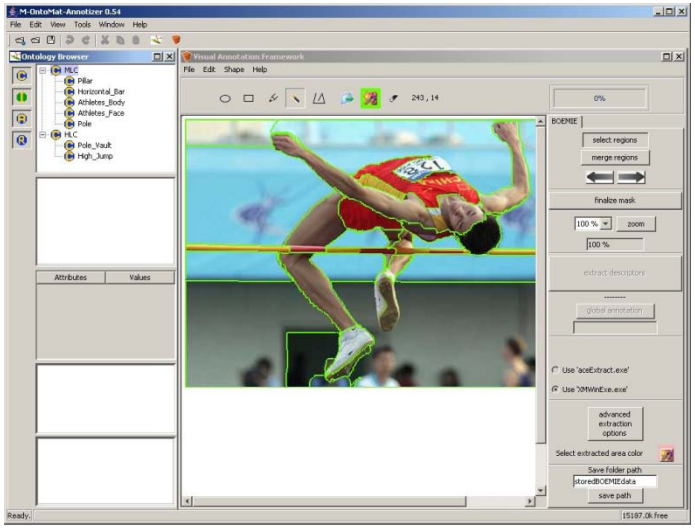
Information Society Technologies

Annotation Tools

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction



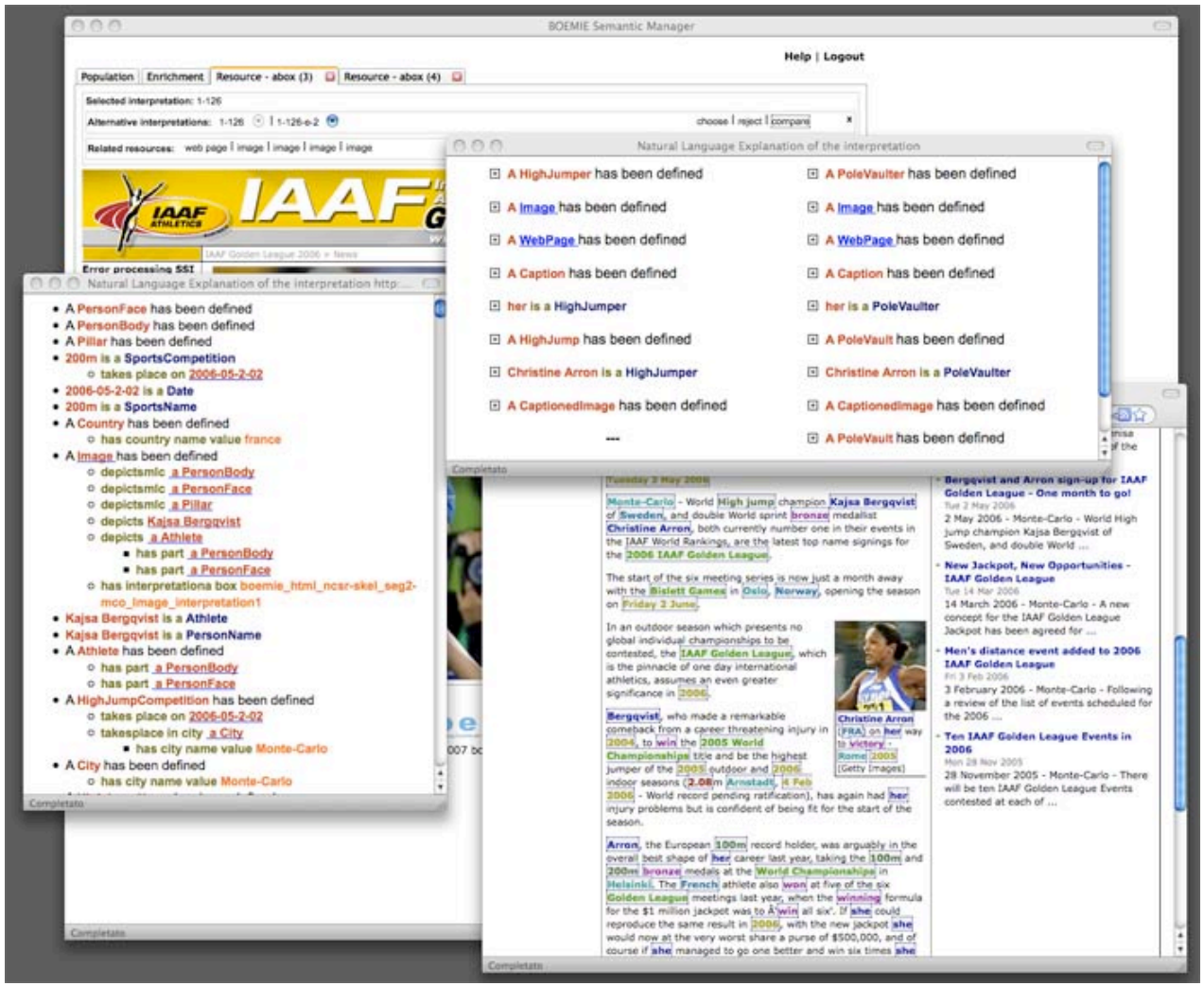




Information Society Technologies

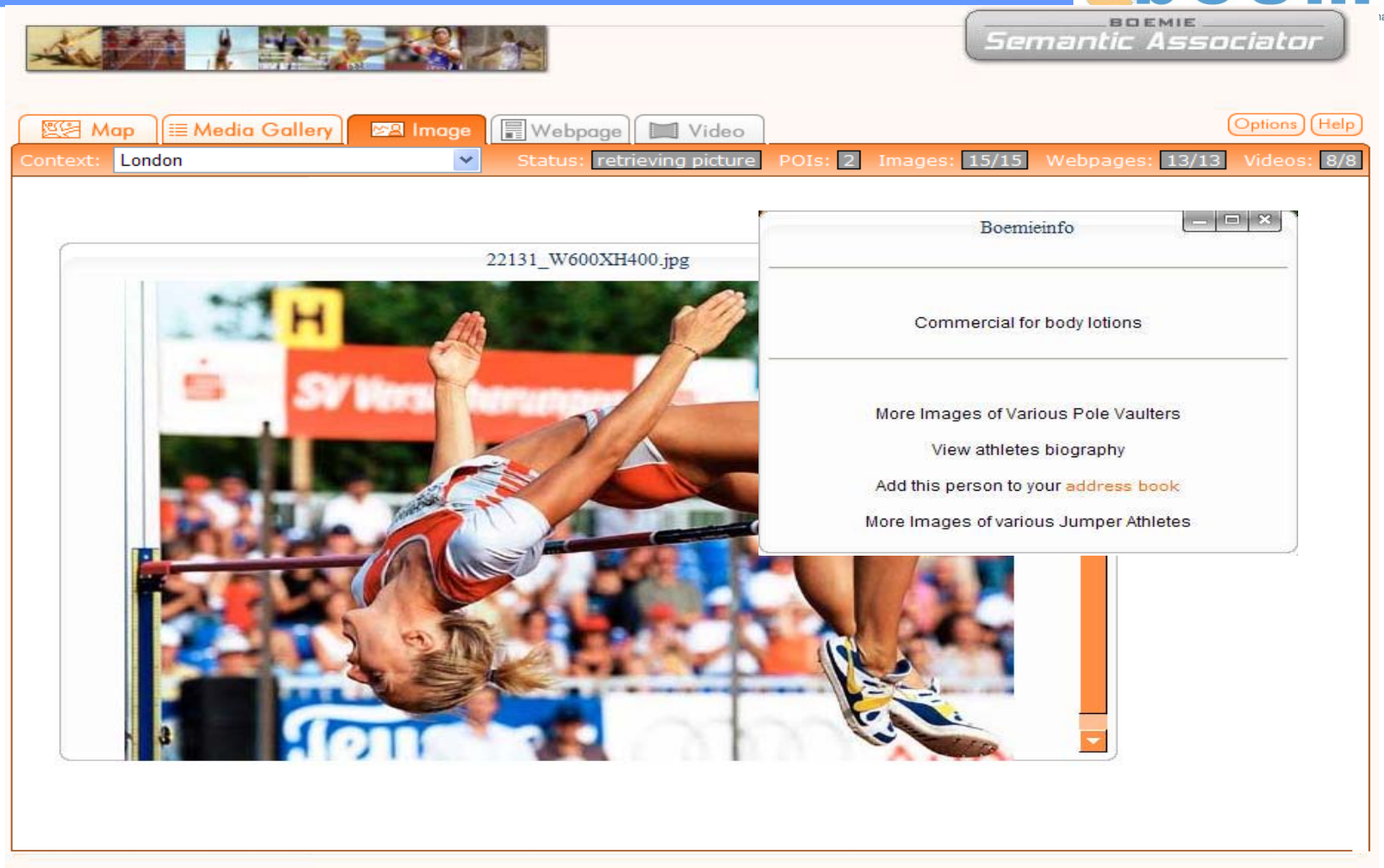
BOEMIE semantic manager

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction



Information Society Technologies

BOEMIE semantic browser



Information Society Technologies

Challenges

- Generic/large-scale recognition in images and video
- Computationally efficient interpretation of extracted information (from mid- to high-level concepts).
- Domain-independent fusion of multiple modalities.
- Hybrid approaches to fused interpretation.
- Discovery of new object types (mid-level concepts).
- Learning new high-level concepts, relations and interpretation rules.
- Human-friendly presentation and definition of new knowledge.

Challenges

- Domain-independent and computationally efficient ontology matching.
- Effective use of external resources in concept definition.
- Intelligent, efficient and user-friendly semantic applications.
- Seamless integration of semantic applications in current working practices (e.g. content production and management).
- Engineering of an efficient and transparent integrated extraction and evolution system that combines many heterogeneous components.

boemie
Bootstrapping Ontology Evolution with Multimedia Information Extraction

http://www.boemie.org/

George Paliouras
(paliourg@iit.demokritos.gr)

Information Society Technologies