

# **EWIMT 2004: Panel on Knowledge-Driven Content Processing and the Semantic Web**

## **Relating natural language semantic to multimedia content features**

**Thierry Declerck: University of Saarland and DFKI GmbH**



Information Society  
Technologies

- Semantic resources as knowledge bases for LT: WordNets (“linguistic ontologies”), Taxonomies (like XBRL for the financial domain), Thesauri (like UMLS for the medical domain) and Ontologies (domain specific or top level).
- All those semantic resources encode their knowledge abstracting over natural language items, terms or expressions (but using natural language for the labelling the abstract items in the knowledge bases - well know problem discussed in philosophy of language).

- Use of Semantic Resources for *Knowledge Markup* of (textual) Web documents -- with authoring tools or supported by automatic text analysis.
- Use of high-level textual analysis for supporting Knowledge Extraction/Learning from (textual) Web Documents.
- A combination of both seems to be appropriate for the implementation of Semantic Web applications
- Some EU projects I am aware of: Esperonto, Sekt ..

# Semantic Web(s) and video/image processing



- Results of video/image analysis are mainly so-called “low-level” features. A semantic gap seems to be existing in the field of content detection (and indexing) in video/image material (keynote speaker yesterday).
- No mono-modal/mono-media analysis (in the present and next period) is being considered as the unique source for a accurate semantic annotation of video/image material. This point mentioned for example by a keynote speaker at CIVR 04.

# Semantic Web(s) and video/image processing



- Need for integration of semantics encoded in associated speech and/or text (superposed text, caption, subtitles etc.) or other available modalities.
- But how to combine both types of “semantic/content” analysis in order to attack the problem of the Semantic Gap but also for supporting the development of a cross-media semantic architecture and applications (like for example the project Direct-Info concerned with Media Monitoring, or the TRECVID challenge)

## Joint annotation initiative on document including both image/video and speech and/or text?



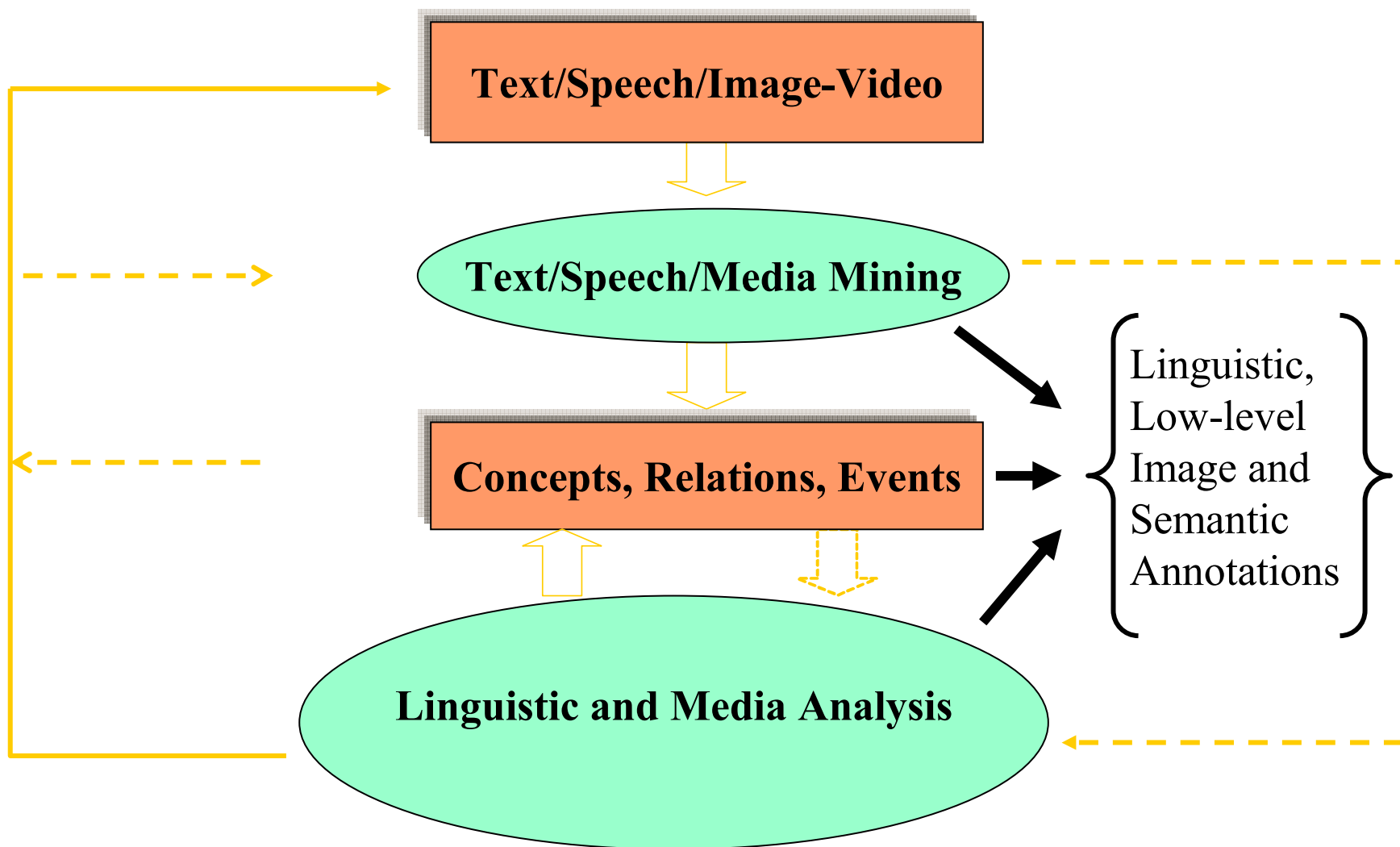
- Compare/Merge annotations provided by image/video and language analysis on documents containing „parallel“ image/video and natural language => training set.
- But this work will be an expensive one in term of human efforts, if appropriate tools are not available.

## Knowledge driven integration/interoperability of key semantic features in both fields



- Work on ontology abstracting over MPEG-7 low-level features already started in aceMedia.
  - Availability of a linguistic description scheme (LDS) in MPEG-7 (a subset of it is used for example in the project Busman and Direct-Info). The LDS is supporting the inclusion of textual information available for (segments of) the image/video annotated with MPEG-7.
- => What about adding an ontology description on the top of the LDS, which shares/integrates concepts over the linguistic and the low-level features?

# The challenge: Cross-Media ontology-based knowledge annotation and extraction/acquisition



Thanks!



- For your attention
- To the organizers of this workshop for inviting me participate at this panel session
- To the coordinator of the SCHEMA network for inviting the Esperonto project to join the network as an affiliated member, and for allowing so to start interesting discussions on those issues