Semantics Extraction from Multimedia Content: The BOEMIE Architecture

Sergios Petridis, Nicolas Tsapatsoulis

***Abstract*— The BOEMIE project is a European Commission research program aiming at ontology evolution through multimedia information extraction. This papers presents an overview of the BOEMIE architecture for semantics extraction from multimedia content.**

***Index Terms*— semantics extraction, multimedia, ontologies, reasoning, multimodal fusion.**

I. Introduction

COMPLEX structured multimedia documents possess a rich variety of information appearing in all sorts of forms and combined under diverse schemata. Their analysis is a demanding operation since a large number of specific per-medium processing techniques need to be developed, assembled and fused in a way that enables their interpretation and adaptation, in the context of a domain application.

In BOEMIE, the purpose of a Methodology for Semantics Extraction from Multimedia Content is to specify how information from the multimedia semantic model can be used to achieve semantic extraction from various modalities (text, image, video and audio) and to come up with an open architecture, which will communicate with the ontology evolution modules in WP4, accessing existing knowlesge and providing back newly extracted information.

II. Architecture

The design choices of the semantics extraction methodology have been guided by the core ontology-oriented architecture of the BOEMIE project. Although ontology is a useful milieu for systematically fusing and interpreting multimedia analysis results, it prompts for devising a particular approach to enable its interfacing with ontology-unaware media processing and machine learning techniques. The architecture for semantics extraction from multimedia content has been designed to advance the state of the art by (a) facilitate independent development of processing and learning techniques per medium (b) allow transparent coordination of per-medium semantics extraction modules and (c) enable reasoning-based feedback on semantics extraction results. Moreover, the architecture supports the evolution of the system, by requiring medium processing algorithms to be adaptable by means of both supervised and unsupervised machine learning techniques.

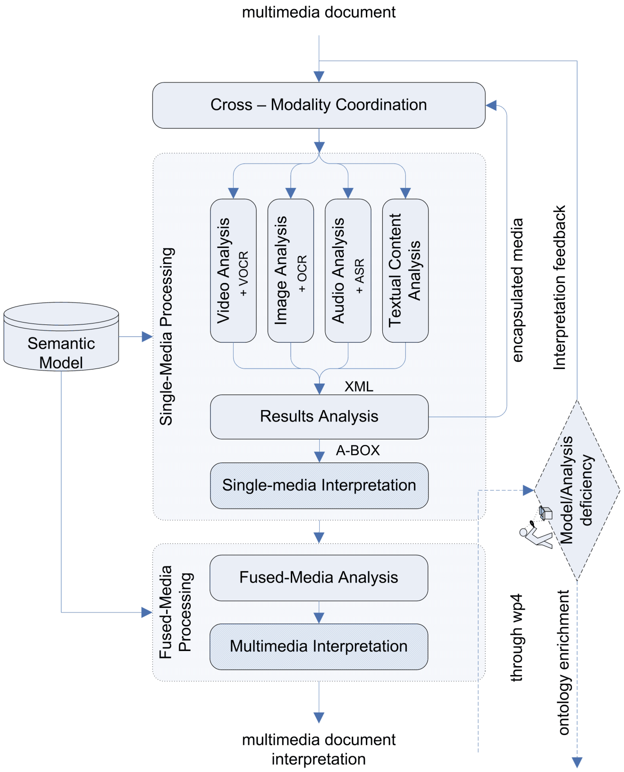

Fig. 1. Processing and interpreting a multimedia document

A. Design principles

The semantics extraction architecture is deployed in two axes which allow, on one hand, to deal with multimodal information and, on the other, to bridge the gap between extraction techniques and ontology-based reasoning services.

- *Multimodal information fusion*: To deal with multimedia documents, semantics extraction is decomposed into two steps. First an analysis of each medium-specific sub-document is done. Then, extraction results are fused, in order to take into account complementarity, redundancy and coherence of the extracted information. These steps may be repeated in a loop to account for (a) analysing embedded documents (such as OCR text in images) and (b) refining the analysis of one medium-specific document using information extracted from an other.
  Multimodal data fusion is explicitly supported by the ontology. Namely, for each modality, a set of modality-specific concepts is defined. Concepts across modalities are then associated with modality-independent concepts

This work was supported by the European Commission under the FP6-027538 contract.

S. Petridis and N. Tsapatsoulis are with the Institute of Informatics and Telecommunications, National Centre for Scientific Research “Demokritos”, Athens, Greece, e-mail: {petridis,ntsap}@iit.demokritos.gr.

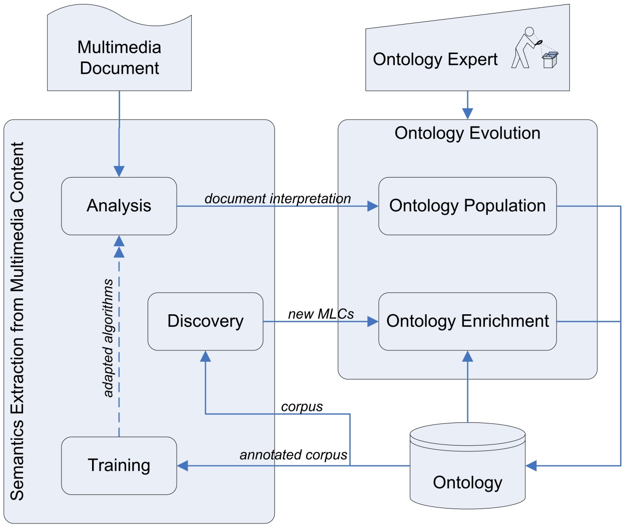

Fig. 2. The bootstrapping process of BOEMIE

(Example: the *visual representation of an athlete*, such as its photo, and the *textual representation of an athlete*, such as its name, are concepts specific to the image and text modality respectively, related to the the modality–independent *athlete* concept).

- *Bridging the semantic gap*: Semantics extraction for a particular medium is further decomposed into two steps. First, segments inside a document are detected and classified using medium-specific processing techniques. The development of specific methodologies to process each medium is a significant part of WP2, separately described in the following section. To allow linkage with the ontology, the set of possible classes per medium are mapped to an evolving subset of concepts of the ontology, referred to as as *mid-level concepts*. (Example: the *visual representation of a pole* is a mid level concept, under the condition that it is possible to detect directly a pole in an image, using the image analysis tools).

  Once mid-level concepts instances in a document have been found, reasoning services, such as deduction and abduction, complement the document analysis, by inferring further existences of instances of aggregate concepts, referred to as *high-level* concepts (Example: a *pole-vault event* is a high level concept, if its existence is deduced by the existence of a *pole* and a *athlete*, together with a suitable rule within the ontology). The findings of reasoning may then be used as a feedback, to refine the detection and classification of mid level concept instances.

B. Semantics extraction adaptability

The semantics extraction methodology comprises three distinct modes of operation, namely *Analysis*, *Training* and *Discovery*. Each mode implements part of the functionality required for BOEMIE to account for ontology population, adaptability of the analysis with respect to new content and direct involvement in ontology enrichment respectively. Although the first mode of operation implements the main semantics extraction task, the second and third modes are essential to the applicability of the BOEMIE system to an evolving domain.

In summary:

- The *Analysis* mode of operation applies each time a new multimedia document becomes available to the BOEMIE system. Its task is to analyse and interpret the document using single-medium specific techniques followed by fusion of multimedia information.
- The *Training* mode of operation applies when new manually annotated content, or content inaccurately analysed so far , is available. Its task is to enhance the analysis modules given the available content, through the usage of supervised machine learning algorithms.
- The *Discovery* mode of operation applies when a significant amount of content is available in the BOEMIE system, that can lead to expansion of the semantic model, through augmenting the set of mid-level concepts. The new concepts correspond to either refinement of existing mid-level concepts or to a clustering of instances so far classified as “unknown”. Clustering of instances is based both on instance similarity and discrimination with respect to a high-level concept which they are associated to.

III. Conclusions and Future Work

The architecture presented in this paper describes the methodology that it will be followed in the framework of the BOEMIE project for semantics extraction from multimedia content. It is expected that by the end of the project the aforementioned architecture will be fully functioning allowing the ontology evolution based on multimedia information extraction to take place. Currently, the text and image analysis submodules have been partially implemented and tested as described in *Analysis*. *Training* has been also utilized for the initialization phase of text and image analysis algorithms (training phase). Audio and video processing modules will be incorporated into *Analysis* soon. The functionality described in *Discovery* mode of operation will be supported once the first two modes have been fixed.

References

[1] BOEMIE: Bootstrapping Ontology Evolution with Multimedia Information Extraction, FP6-027538, 2006-2009, (online at: http://www.boemie.org).

[2] BOEMIE Deliverable D2.1: “Methodology for Semantics Extraction from Multimedia Content”, November 2006.