

Demonstration of the EUMSSI Platform for Multimodal Analysis and Annotation

Jens Grivolla, Maite Melero
Computational Linguistics Group
Universitat Pompeu Fabra
Barcelona, Spain
e-mail: {jens.grivolla, maite.melero}@upf.edu

Yannick Estève
Université du Maine
Le Mans, France
email: yannick.esteve@lium.univ-lemans.fr

Eelco Herder
L3S Research Center
Hannover, Germany
email: herder@l3s.de

Peggy van der Kreeft
Deutsche Welle
Bonn, Germany
email: peggy.van-der-kreeft@dw.de

Raul Marin
VSN Innovations and Media Solutions
Alicante, Spain
email: rmarin@vsn.es

Jean-Marc Odobez
IDIAP Research Institute
Martigny, Switzerland
email: odobez@idiap.ch

Susanne Preuß
Gesellschaft zur Förderung der
Angewandten Informationsforschung
Saarbrücken, Germany
email: susannep@iai.uni-sb.de

Abstract— The project EUMSSI is developing a multimodal platform for identifying and aggregating data presented as unstructured information in sources of very different nature including text, video, audio and social media, thanks to the integration of state-of-the-art text and A/V analysis technologies, semantic enrichment and reasoning, social intelligence and collaborative content-based recommendation. Two demonstrators are being built on top of the platform, with the purpose of showcasing its exploitation potential.

Keywords- multimodal analytics; data mining; semantic web.

I. INTRODUCTION

We present the ongoing FP7-ICT-2013-10 project EUMSSI (Event Understanding through Multimodal Social Stream Interpretation) [1], which develops technologies for identifying and aggregating data presented as unstructured information in sources of very different nature (text, video, audio and social media), including media from online news providers and broadcasters, as well as from social sites (e.g., Twitter, YouTube).

This is accomplished thanks to the integration in a cross-modal semantic representation framework of state-of-the-art technologies from the following fields: (i) text analysis, (ii) audio-visual (A/V) analysis, (iii) semantic enrichment, (iv) reasoning, (v) social intelligence, and (vi) collaborative content-based recommendation.

The multimodal interpretation EUMSSI platform, in an optimized process chain, analyses a vast amount of multimedia content and enriches it with additional metadata layers. These metadata layers provide a multi-faceted access of comprehensive information and can be used for further analyses or enrichment. A core idea is that the aggregation of information is carried out in an interactive manner, so that the metadata resulting from one media helps reinforce the analysis results from other media.

The resulting multimodal, interoperable platform, through the use of innovative visualization techniques will allow an interactive exploration of the enriched data and the discovery of new, implicit information.

The EUMSSI platform will be potentially useful for any application in need of automatic cross-media data analysis and interpretation, such as intelligent content management, recommendation, real time event tracking, content filtering, etc.

II. MULTIMODAL ANALYTICS AND SEMANTIC ENRICHMENT

For interpreting and reasoning over the multimedia data, the EUMSSI platform needs to identify topics, to recognize actors and places, to assign temporal information, and to analyze opinion, among other things. A core idea is that the process of integrating information coming from the different media sources is carried out in an interactive manner, so that the analysis results get mutually reinforced. For example, the

name of a person appearing as a caption in a video is captured thanks to Optical Character Recognition (OCR) techniques; it is then identified as being a person's name by Text Analysis; and it is finally used by the audio analysis to reinforce speaker recognition.

Once the entities and concepts have been identified in the different modalities, all the information is aggregated and semantically enriched, using ontology-based reasoning mechanisms.

III. ALIGNED DATA REPRESENTATION AND FLOW MANAGEMENT

The results of the different analysis components, as well as the original metadata, are normalized to a common metadata schema (based on the *schema.org* ontology) and then stored in the Common Analysis Structure (CAS) format [2] provided by the Apache Unstructured Information Management Architecture (UIMA) [3]. This format allows for an aligned view of the different annotation layers, which is necessary for the cross-modal integration.

Annotations are stored "stand-off", meaning that the original content is not modified in any way by adding annotations. Rather, the annotations are entirely separate and reference the original textual content by offsets. In the case of multimedia content (video and audio) annotations will naturally refer to that content via timestamps, which are the common base for alignment between the different annotation layers. Hence, any textual views created from multimedia content, be it a text automatically recognized from the audio, or extracted from the video by OCR, refer back to the timestamps in the original content.

All text analysis components, as well as the integration and reasoning components, are implemented as internal UIMA analysis engines, while audio and image components are external.

All the data and metadata are stored in a MongoDB database [4]. Precisely, the design of the flow management, rather than being procedural, is presented in terms of transformations between data states. Thus, in order to avoid synchronization issues, the state of the data processing is asynchronously stored together with the data. The list of pending tasks can be extracted at any point through a simple database query. The resulting system is more consistent and robust, relying only on the data states to protect itself against potential server failures.

The aggregated semantic information is finally mirrored to a Solr search engine [5], where it is indexed and can be queried by the relevant applications.

IV. EXPECTED FINAL RESULTS AND POTENTIAL IMPACT

On the one hand, at a primary level, the EUMSSI multimodal-based metadata automatic enrichment capabilities can complement any Multimedia Asset Management (MAM) system.

On the other hand, the project is building two demonstrators on top of the platform aimed at showing its

exploitation potential: (i) a Computer Assisted Storytelling tool, and (ii) a Second-screen application.

(i) Through the Computational Assisted Storytelling demonstrator, we expect EUMSSI to boost productivity of the overall news production workflow (such as the one used by Deutsche Welle's journalists), by providing an efficient manner to automate and integrate monitoring, gathering and filtering tasks in the news article creation lifecycle. Hence, while the journalist is editing a story, he will automatically be presented with a series of related content, suggestions, hot topics, from many sources ranked in turn by relevance, up-to-dateness, reliability, category, etc., by means of an array of visualization devices that will help him explore the data in an interactive and principled way, and discover hidden relevant information in large document collections.

(ii) With its Second-screen demonstrator, EUMSSI has the potential to enhance the media watching experience of the European citizens by providing them with personalized visualizations and games, aimed to further engage them and to allow them to investigate background information and related aspects, as well as share their findings in social media.

All technical results of the project will be accessible to the community, both research and industrial, by licensing them using the Meta-Share platform, under an open source licensing scheme, allowing a commercial exploitation of the developed resources and tools and also an unrestricted use for research purposes.

NOTE

All developments are publicly available at [6].

More information about the project can be found at the official EUMSSI website [7].

ACKNOWLEDGMENT

The work presented here is being carried out within the FP7-ICT-2013-10 STREP project EUMSSI under grant agreement n° 611057, receiving funding from the European Union's Seventh Framework Programme managed by the REA (Research Executive Agency) [8].

REFERENCES

- [1] J. Grivolla, M. Melero, T. Badia, C. Cabulea, Y. Esteve, E. Herder, J.-M. Odobez, S. Preuss, and R. Marín, "EUMSSI: a Platform for Multimodal Analysis and Recommendation using UIMA," Proc. of the Workshop on Open Infrastructures and Analysis Frameworks for HLT (COLING), pp. 101--109, Dublin, Ireland, August, 2014
- [2] Ferrucci D. et al. (2006) Towards an interoperability standard for text and multi-modal analytics. IBM Res. Rep., RC2
- [3] <http://uima.apache.org/>
- [4] <http://www.mongodb.org/>
- [5] <http://lucene.apache.org/solr/>
- [6] <https://github.com/EUMSSI/>
- [7] <http://www.eumssi.eu/>
- [8] <http://ec.europa.eu/research/rea>