

MULTISENSOR: DEVELOPMENT OF MULTIMEDIA CONTENT INTEGRATION TECHNOLOGIES FOR JOURNALISM, MEDIA MONITORING AND INTERNATIONAL EXPORTING DECISION SUPPORT

Stefanos Vrochidis¹, Ioannis Kompatsiaris¹, Gerard Casamayor², Ioannis Arapakis³, Reinhard Busch⁴, Vladimir Alexiev⁵, Emmanuel Jamin⁶, Michael Jugov⁷, Nicolaus Heise⁸, Teresa Forrellat⁹, Dimitris Liparas¹, Leo Wanner^{2,10}, Iris Miliaraki³, Vera Aleksic⁴, Kiril Simov⁵, Alan Mas Soro⁶, Mirja Eckhoff⁷, Tilman Wagner⁸, Martí Puigbó⁹

¹Centre for Research and Technology Hellas, ²Pompeu Fabra University, ³Barcelona Media-Yahoo! Labs, ⁴Linguatrec, ⁵Ontotext, ⁶everis, ⁷pressrelations, ⁸Deutsche Welle, ⁹PIMEC, ¹⁰ICREA stefanos@iti.gr

ABSTRACT

This paper presents an overview and the first results of the FP7 MULTISENSOR project, which deals with multidimensional content integration of multimedia content for intelligent sentiment enriched and context oriented interpretation. MULTISENSOR aims at providing unified access to multilingual, multimedia and multicultural economic, news story material across borders in order to support journalism and media monitoring tasks and provide decision support for internationalisation of companies.

Index Terms — Content integration, sentiment, summarisation, multimedia, social media, mining, context

1. INTRODUCTION

During the past decade, the rapid development of digital technologies and the low cost of recording media have led to a great increase in the availability of multilingual and multimedia content worldwide. In the best case, this content is repetitive or complementary across political, cultural, or linguistic borders. However, the reality shows that it is also often contradictive and in some cases unreliable. The consumption of such large amounts of content regardless of its reliability and cross-validation can have important consequences on the society and especially on journalism, media monitoring and international investments.

In order to break this isolation, there is a need for technologies that are capable of capturing, interpreting and relating economic information and news from various subjective views as disseminated via TV, radio, newspapers, blogs and social media. MULTISENSOR aims at bridging this gap by envisaging at an integrated view of heterogeneous resources sensing the world (i.e. sensors), such as international TV, newspapers, radio and social media. The approach of MULTISENSOR builds upon the

concept of multidimensional content integration (Fig. 1) by considering the following dimensions for mining, linking, understanding and summarising heterogeneous material: language, multimedia, semantics, context, emotion, time and location. Thus, the overall goal of MULTISENSOR is to research and develop a unified platform, which will allow for the multidimensional content integration from heterogeneous sensors, with a view to providing end-user services such as journalism and media monitoring, and decision support for SME internationalisation.

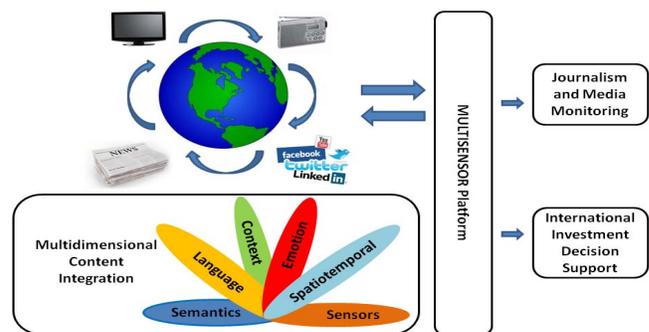


Fig. 1. The MULTISENSOR concept

Towards the development of a unified platform that allows for multidimensional content integration considering the aforementioned dimensions, MULTISENSOR addresses the following *scientific objectives*:

- Mining and content distillation of unstructured heterogeneous and distributed multimedia and multilingual data, including semantic concept extraction.
- User- and context-centric analysis of multimedia and multilingual content. This objective aims at analysing content from the user perspective to extract sentiment and context, as well as analysing computer-mediated interaction in the web and specifically in social media.

- Semantic integration of heterogeneous multimedia and multilingual data.
- Development of hybrid reasoning techniques for intelligent decision support.
- Context-aware multimodal aggregation and multilingual summarisation, as well as adequate presentation of the information to the user.

2. USER PERSPECTIVE

In order to ensure that the proposed work is guided by a user-centric view and in order to grant general usability and exploitability of the project results, three pilot *use cases* (UC) have been defined and specific requirements have been extracted for each of them:

UC1: Journalism: Journalists need to master large heterogeneous amounts of multimedia and multilingual data when writing a new article. On the basis of a market analysis that was conducted and from a journalistic point of view, MULTISENSOR should be able to provide an automatic summarisation of heterogeneous and multilingual digital information. The platform should also automatically suggest related content and information that allows journalists to enrich their coverage of a specific topic.

UC2: Commercial media monitoring: Professional clients of media monitoring portals require direct access to comprehensive and targeted business and consumer information. This could include information on consumption habits, competitors and opinions. From a media monitoring point of view, it is important that the MULTISENSOR system follows the usual workflow for the creation of a media analysis. In a first step, the user needs to define the sources and time frame that is to be monitored, along with the search terms he wants to use. In a second step, the search results need to be curated and validated. The MULTISENSOR system should present the results of these queries in different output formats and visualisations.

UC3: SME (Small and Medium Enterprises) internationalisation: This UC deals with SME internationalisation, which refers to small or medium-sized companies that want to start or are in the process of expanding from a regional or a national market to a new and foreign market in order to increase turnover and profit. This process is of particular importance as it is often the only option to achieve growth. But it is also aligned with considerable challenges such as a lack in knowledge about market conditions or the spoken language in the targeted countries. From the aforementioned, in order for the MULTISENSOR platform to be fully helpful in SME internationalisation cases and improve the decision-making process, it should provide information about several related indicators, regarding the condition of the market, the political and financial situation of the countries, potential competitors, consumption habits, etc. Furthermore, two very important requirements from this UC are summarisation (to reduce the amount of information that the

internationalisation expert will need to read and study) and automatic language detection and translation.

All three UC reflect the challenge of having to deal with a large amount of heterogeneous data and information from many different sources and in many different languages. However, there are also very significant differences. For example, journalists and commercial media monitors are interested in continuously tracking a specific topic, brand or campaign. The ultimate goal is to update their audience or clients on a regular basis. Although this information will ultimately lead to specific decisions, the priority is information gathering and analysis.

On the other hand, the SME internationalisation UC focuses mainly on decision support. In order to make a decision about whether a company should move into a new market, owners and managers need to know about all relevant market indicators, opportunities and barriers. This includes information about consumption, competition, socio-economic indicators, as well as information about legal restrictions or other statutory requirements.

The functionality of the MULTISENSOR framework considers all the requirements imposed by the three UC.

3. MULTISENSOR FRAMEWORK

The architecture of the MULTISENSOR framework is depicted in Fig. 2. In this architecture, a periodic process of content harvesting takes place, which retrieves source material by crawling a set of sources for news, multimedia and social network content. Next, the different components of the framework, as well as the functionality of the modules that they contain and provide are described.

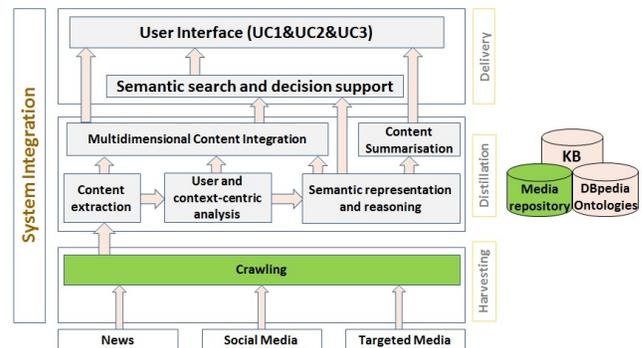


Fig. 2. Architecture of the MULTISENSOR framework

3.1. Multimedia content extraction

This component aims at extracting knowledge from multimedia input data and presenting the extracted knowledge in a way that subsequent components can operate on it. It includes the following technologies:

1) **Language Identification:** Before a text is stored in the repository, it is analysed in which language it is written and the text is annotated accordingly. The languages considered

in MULTISENSOR are English, German, Spanish, Bulgarian and French.

2) **Named entities extraction:** This module aims at identifying *names* (named entities) in texts. Names are words which uniquely identify objects, like ‘Berlin’, ‘Siemens’, etc. The module incorporates two linguistic components that allow all analysis modules to operate on the same input: sentence segmentation and tokenisation.

3) **Concept extraction from text:** Concept extraction starts from the results of the named entities extraction task. The goal of this module is to identify in the text mentions to concepts that belong to the project domains. Candidate concepts are identified through analysis of multilingual corpora. When processing new documents, the module attempts disambiguation of mentions of concepts against relevant ontologies and datasets.

4) **Concept linking and relations:** This module aims at identifying in texts relations between mentions of named entities and concepts. Two relation types are considered: i) coreference relations i.e. several mentions make reference to the same entity, and ii) n-ary relations describing situations and events involving multiple entities and concepts. To this end, a deep dependency parser [1] that delivers deep-syntactic dependency structures from sentences in nature language has been developed. This parser uses the output of an optimised dependency parser [2] as input.

5) **Audio recognition and analysis:** Automatic speech recognition (ASR) is employed in order to provide a channel for analysis of spoken language in audio and video files. The transcripts produced follow the same analysis procedure as the input from other text sources. The languages covered by the ASR component are English and German.

6) **Multimedia concept and event detection:** This module receives as input a multimedia file (i.e. image or video) and computes degrees of confidence for a predefined set of visual concepts. The module performs video decoding (applicable for video files only), feature extraction and classification in order to assign a confidence value for a concept or event existence in an image or video shot [3].

7) **Machine translation:** Automatic machine translation (MT) has two main goals: to provide the translation of the summarisation results in the end of the content analysis and summarisation chain and to enable full-text translation on-demand during the development of language dependent analysis tools in the project, in case a subset of required languages is not supported by these tools.

3.2. User and context-centric analysis

The objectives of this component are to model and represent contextual, sentiment and online social interaction features, as well as deploy linguistic processing at different levels of accuracy and completeness.

1) **Extraction of contextual features:** This module provides a set of contextual indicators characterising the content items and a framework for measuring their impact in the

context of the UC. Moreover, it provides representation techniques to be used in effective context-based search.

2) **Polarity and sentiment extraction:** The polarity and sentiment extraction module aims at modelling a robust opinion mining system that is based on linguistic analysis and is applicable to large datasets. Moreover, models that take into account the presence of named entities in different sentences have been designed within the module.

3) **Contributor analysis:** This module deals with online social interaction. Its functionality focuses on analysing complex networks in order to retrieve the social interactions and the social profile of a specific contributor (e.g. author).

3.3. Multidimensional content integration and retrieval

The objective of this component is to achieve integration and retrieval of content along different dimensions.

1) **Multimodal indexing and retrieval:** In this module, a multimedia data representation framework that allows for the efficient storage and retrieval of socially connected multimedia objects is developed. The representation model is called *SIMMO* (Socially Interconnected MultiMedia-enriched Objects) [4] and has the ability to fully capture all the content information of interconnected multimedia objects, while at the same time avoiding the complexity of previously proposed models.

2) **Topic-based modelling:** In this module, two subtasks are considered: a) category-based classification and b) topic-event detection. The module receives as input multimodal features that are created in the multimedia content extraction component and provides as output the degree of confidence of a number of categories for a specific content item (for category-based classification) or a grouping for a list of content items based on the existence or not of a number of topics / events (for topic-event detection).

3.4. Semantic representation and reasoning

MULTISENSOR includes a semantic layer in order to represent in a unified way heterogeneous content. The following technologies are involved:

1) **Semantic representation:** This representation includes a number of ontologies that are integrated in a common framework, such as DBpedia, GeoNames and FreeBase.

2) **Ontology alignment:** The ontology alignment module discovers candidate semantic correspondences between heterogeneous information descriptions and terminologies and verifies the correctness and consistency of the discovered mappings in an automatic way.

3) **Content alignment:** This module deals with the semantic processing of the multimodal content, in order to identify near duplicate and contradictory information relying on semantic technologies.

4) **Hybrid reasoning and decision support:** In the hybrid reasoning and decision support module, four reasoning techniques are developed: hybrid reasoning (consists of a

combination of forward and backward chaining), multi-threaded reasoning (allows parallel inference calculation), temporal reasoning (ensures inference based on temporal entities and sequence in time) and geo-spatial reasoning (provides the ability to reason based on latitude, longitude and altitude of a given location). Additionally, a reasoning-based recommendation system with two main functionalities is developed: firstly, it determines relevant facts by navigating the graph and secondly, it advises the user by interpreting these facts through the use of the aforementioned hybrid reasoning techniques and the assignment of relevance weights for each selected fact.

3.5. Content summarisation

The content summarisation component implements procedures for producing multilingual briefings. Two established strategies in the field of text summarization are considered in MULTISENSOR:

1) **Extractive summarisation:** Text-to-text summarization, where the relevance of sentences in the original documents is assessed based on shallow linguistic features in order to decide on its inclusion of a summary. A module following this strategy is used in order to establish a basic infrastructure for summarisation services and implement a fall-back method.

2) **Abstractive summarisation:** Documents are analysed and the information extracted from them is used to generate a summary that is not composed of fragments of the original documents, but is generated directly from data. A module implementing abstractive methods operates on the semantic layer in order to select contents extracted from multimedia documents and also coming from other datasets integrated into the MULTISENSOR system. Contents are selected and organized into a text plan that guarantees the coherent presentation of information. This is achieved by using models derived from corpora of texts annotated with data (e.g. topic, entities and concepts, sentiment analysis). These corpora are produced by the multimedia content extraction pipelines by enriching texts with annotations of the contents extracted from them. A multilingual linguistic generation system renders text plans into the final summaries.

3.6. Final platform

The aforementioned technologies are integrated in a service-oriented platform. A screenshot of the MULTISENSOR platform can be seen in Fig. 3.

4. CONCLUSIONS AND FUTURE WORK

During the first 1,5 year of its lifetime, MULTISENSOR has achieved considerable progress regarding all of the objectives mentioned in Section 1. Specifically, the first MULTISENSOR prototype integrating the initial versions of the research modules has been implemented.

In the remaining 1,5 year, MULTISENSOR will carry on the research and development tasks related to the individual objectives and the accomplishment of the corresponding milestones. More specifically, for the remainder of this year MULTISENSOR will move towards the implementation of the second MULTISENSOR prototype, which will integrate the advanced versions of the research modules. Finally, during the third and final year, the project will focus on the development of the final platform and its evaluation.

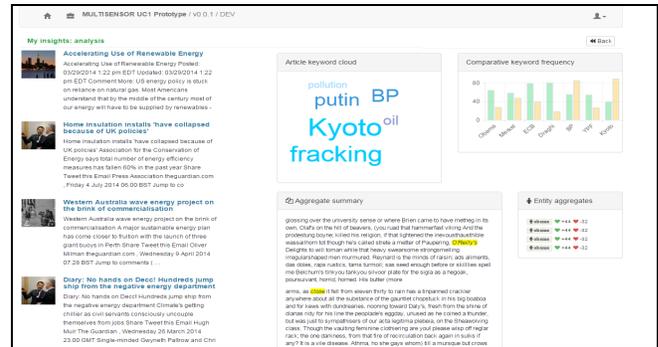


Fig. 3. The MULTISENSOR platform

5. ACKNOWLEDGEMENTS

This work was supported by MULTISENSOR project¹, partially funded by the European Commission, under the contract number FP7-610411.

6. REFERENCES

[1] M. Ballesteros, B. Bohnet, S. Mille and L. Wanner, “Deep-Syntactic Parsing,” COLING 2014, Dublin, Ireland, 2014.

[2] M. Ballesteros and B. Bohnet, “Automatic Feature Selection for Agenda-Based Dependency Parsing,” COLING 2014, Dublin, Ireland, 2014.

[3] N. Gkalelis, F. Markatopoulou, A. Moutzidou, D. Galanopoulos, K. Avgerinakis, N. Pittaras, S. Vrochidis, V. Mezaris, I. Kompatsiaris and I. Patras, “ITI-CERTH participation to TRECVID 2014,” Proc. TRECVID 2014 Workshop, Orlando, FL, USA, Nov. 2014.

[4] T. Tsirikia, K. Andreadou, A. Moutzidou, E. Schinas, S. Papadopoulos, S. Vrochidis and I. Kompatsiaris, “A Unified Model for Socially Interconnected Multimedia-Enriched Objects,” In *MultiMedia Modeling*, pp. 372-384, Springer International Publishing, Jan. 2015.

¹ <http://multisensorproject.eu/>